ALBERT-LUDWIGS-
UNIVERSITÄT FREIBURG

Algorithms for Radio Networks

Data Aggregation

University of Freiburg
Technical Faculty
Computer Networks and Telematics
Christian Schindelhauer

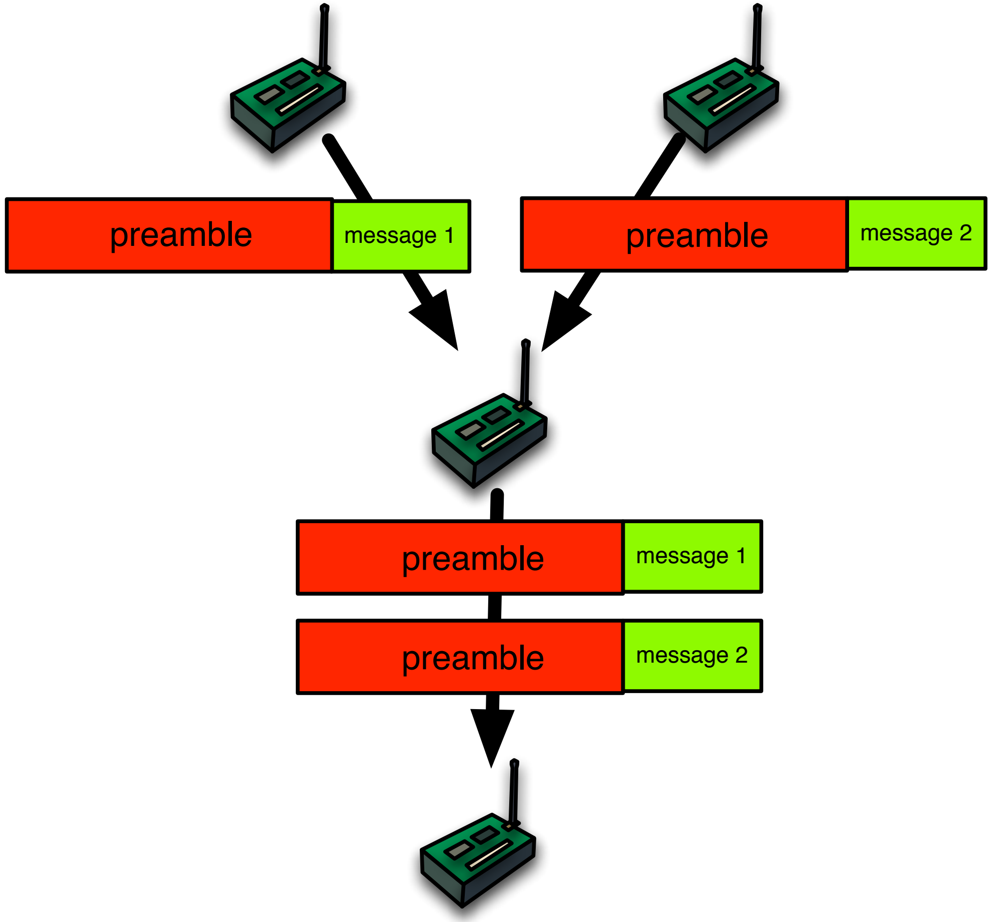
CoNe Freiburg

IIF
INSTITUT FÜR
INFORMATIK
FREIBURG

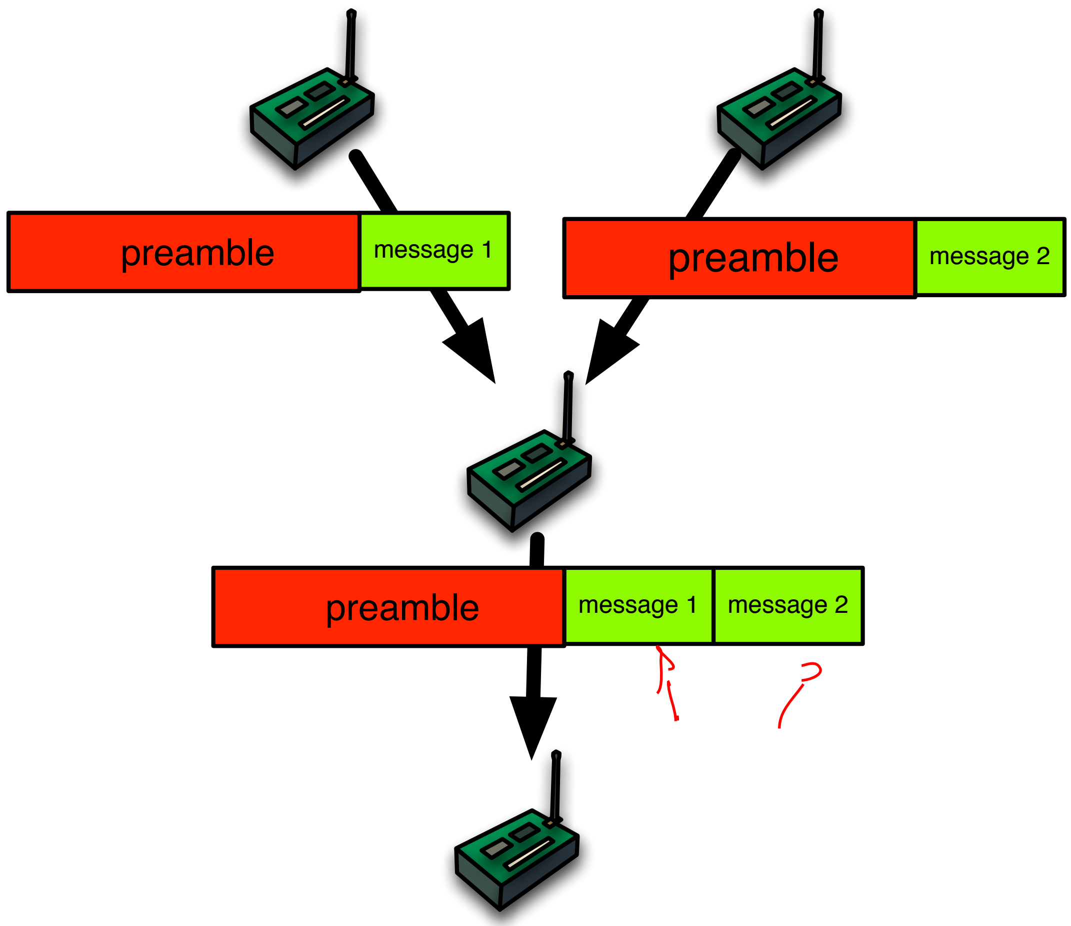
Data Aggregation

- **In multi-hop networks combining mesage can improve networking**
- **Concatenation of messages**
  - overall number of headers is reduced
    - especially for Preamble Sampling
  - smaller costs for collision avoidance
- **Recalculation of contents**
  - e.g. If the minimum temperature is required, then it satisfies to forward the smallest value
  - For this purpose, collect the input over some time

No Data Aggregation



Data Aggregation by Concatenation



Real Data Aggregation by Recalculation

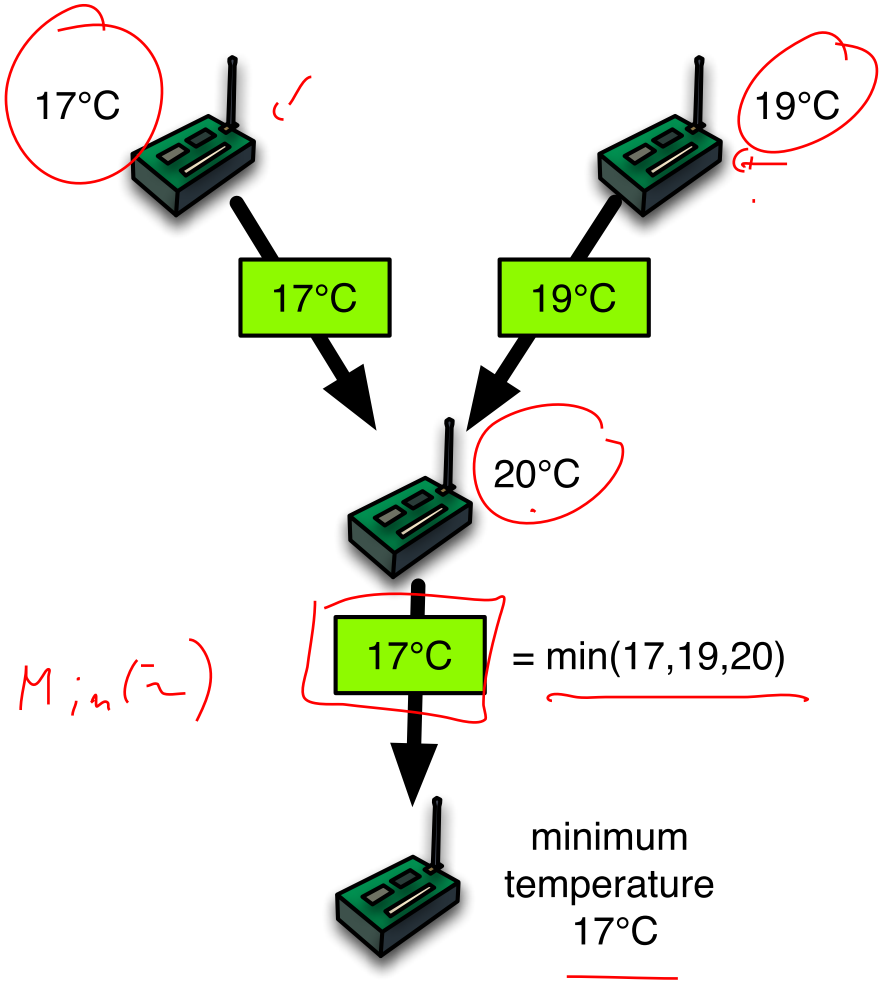
Min, Average ✓

Max, Σ,

Counting,

$$\sqrt{\frac{a^2+b^2}{2}}$$

$$\mathrm{Max}(x) = -\mathrm{Min}(-x)$$



Simple Functions for Data Aggregation

- **Minimum**
  - inner node computes the minimum of input values
- **Maximum**
  - like Minimum
- **Number of sources**
  - inner node adds input values
- **Sum**
  - addition at inner nodes

Aggregable Functions

- **Mean**
  - compute the number of sensors: n
  - compute the sum of sensor values: S
  - mean = S/n
- **Variance**
  - Compute average and the average of squares of values
  - $V(X) = E(X^2)-E(X)^2$

$\sqrt{V(X)}$

$\Sigma x_i \quad \Sigma x_i^2 \quad n$

$$\sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2}$$

Non-Aggregable Functions

- **The following functions cannot be aggregated easily**
  - median
  - p-quantile
    - if p is not very small or large
  - number of different values
    - only for large data sets an approximation is possible
- **Approximate solution**
  - was presented in „Medians and Beyond: New Aggregation Techniques for Sensor Networks, Shrivastava et al. Sensys 04
  - using k words in each message an approximation ratio of ~~log~~ (log n)/k can be achieved

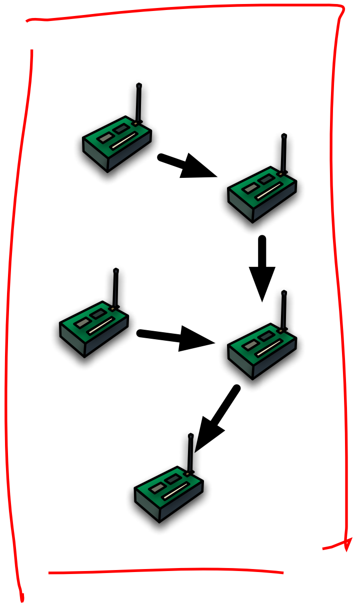
Routing Models for Data Aggregation

- **Address Centric Protocol**
  - each sensor sends independently towards the sink
  - not suitable for (real) aggregation
- **Data Centric Protocol**
  - Forwarding nodes can read and change messages

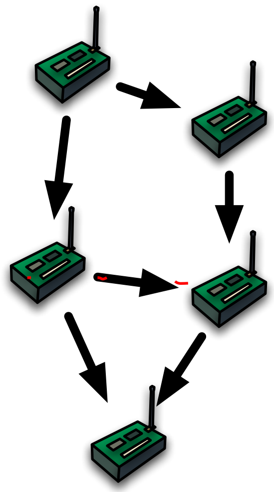
Communication Graphs

- **Tree Structure**

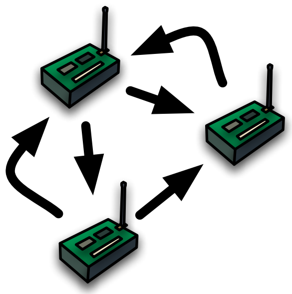
  Tree

  - If there is only a single sink
  - and every source uses only a single path
  - then every communication graph in a WSN is a tree
- **DAG (directed acyclic graph)**

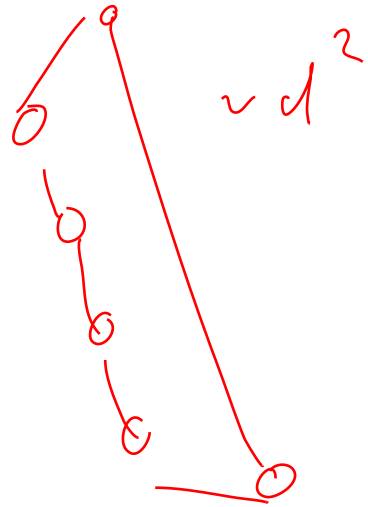
  - general case
  - caused by changing routing paths to the sink
  - may complicate data aggregation
    - e.g. sum
- **General graph**

  - Population protocols
  - are not used in WSNs

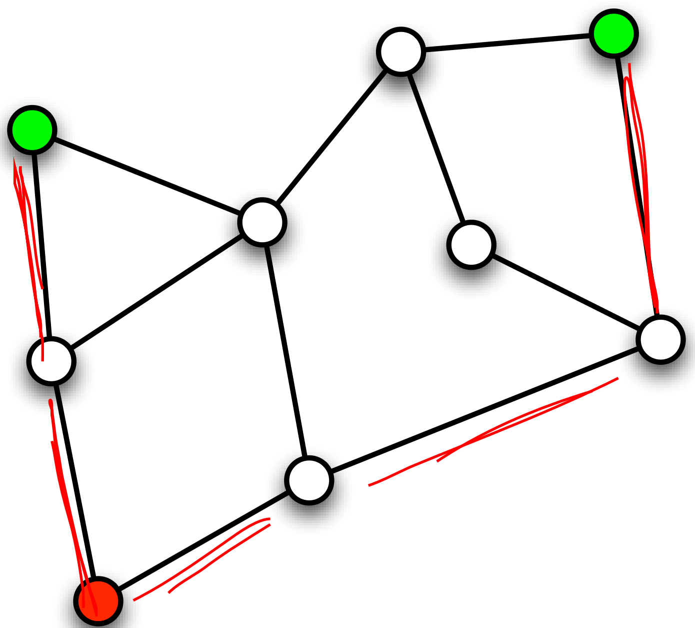
Energy Optimal Tree Structure

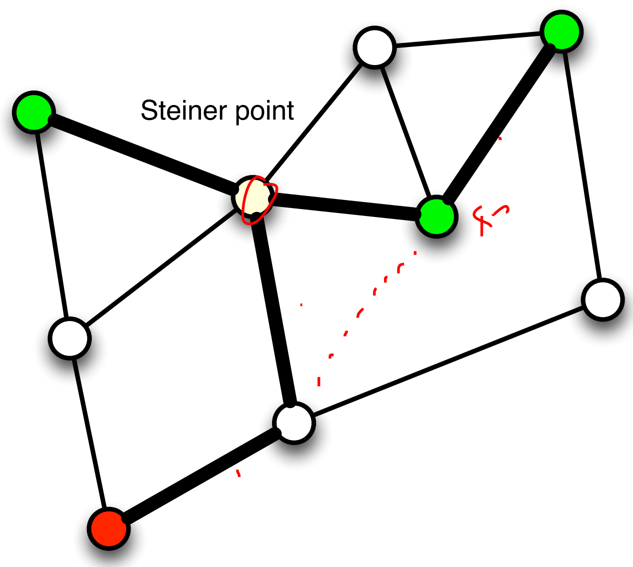
- **Given:**
  - set of data sources and a sink
  - communication graph G
- **Compute:**
  - Steiner tree T
    - sub-graph of G
    - connects all sources and sinks
    - number of edges is minimal
- **Alternative:**
  - edges have an (energy) weight
  - minimize the sum of edge weights



Steiner Tree Problem

- **Observation:**
  - sources and sinks can be handled the same way





Optimal Choice of Data Aggregation

- **Steiner tree**
  - minimal tree connecting all sources and sinks
- **Computation of the Steiner tree is NP-hard**
- **Approximation**
  - the Steiner tree problem can be approximated with factor 2
    - if the underlying graph is metric
  - best known approximation factor for algorithms in polynomial time: 1.55
    - Zelikovsky, Robins 2006

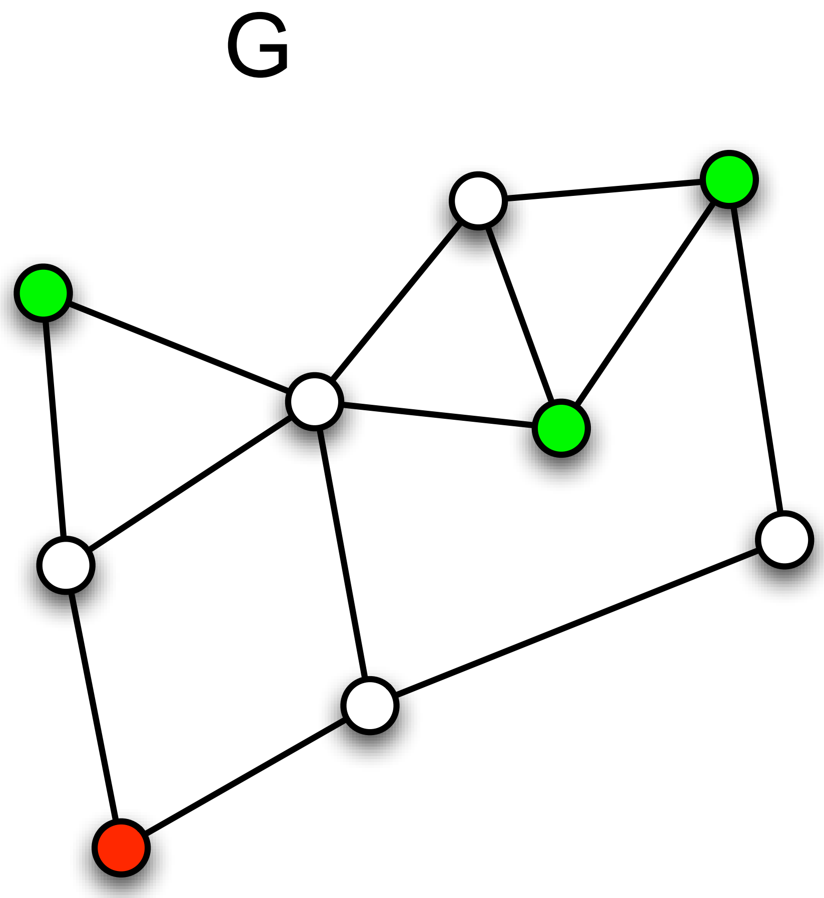
Approximation with the Help of MST

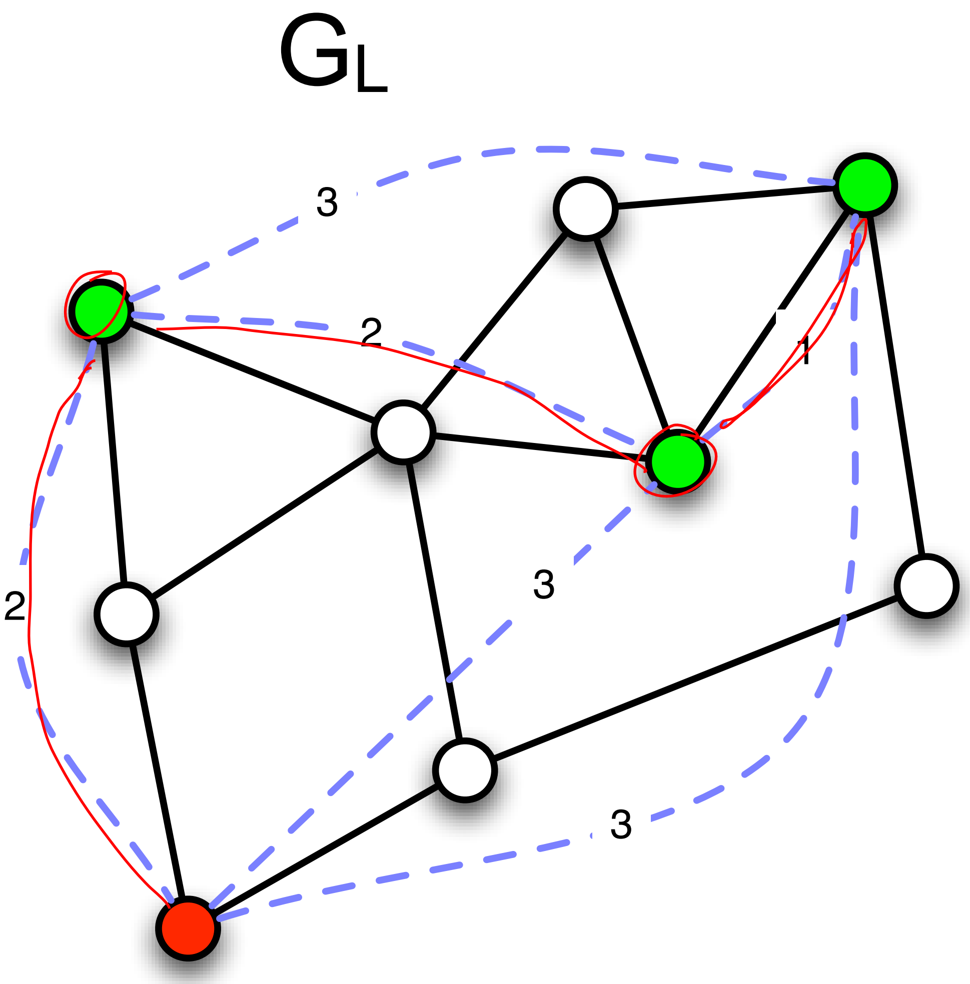
- Source:
  - Information Processing Letters, 27 (1988), 125-128, Kurt Mehlhorn, *A Faster Approximation of the Steiner Tree Problem*
- Compute the distance between the terminal nodes in the graph G
  - Compute the complete graph $\mathbf{G_L}$ with terminal nodes V_T and edge weight according to the distance in G
- Compute **minimum spanning tree** in G_L
- Initialize tree T with empty set

network information

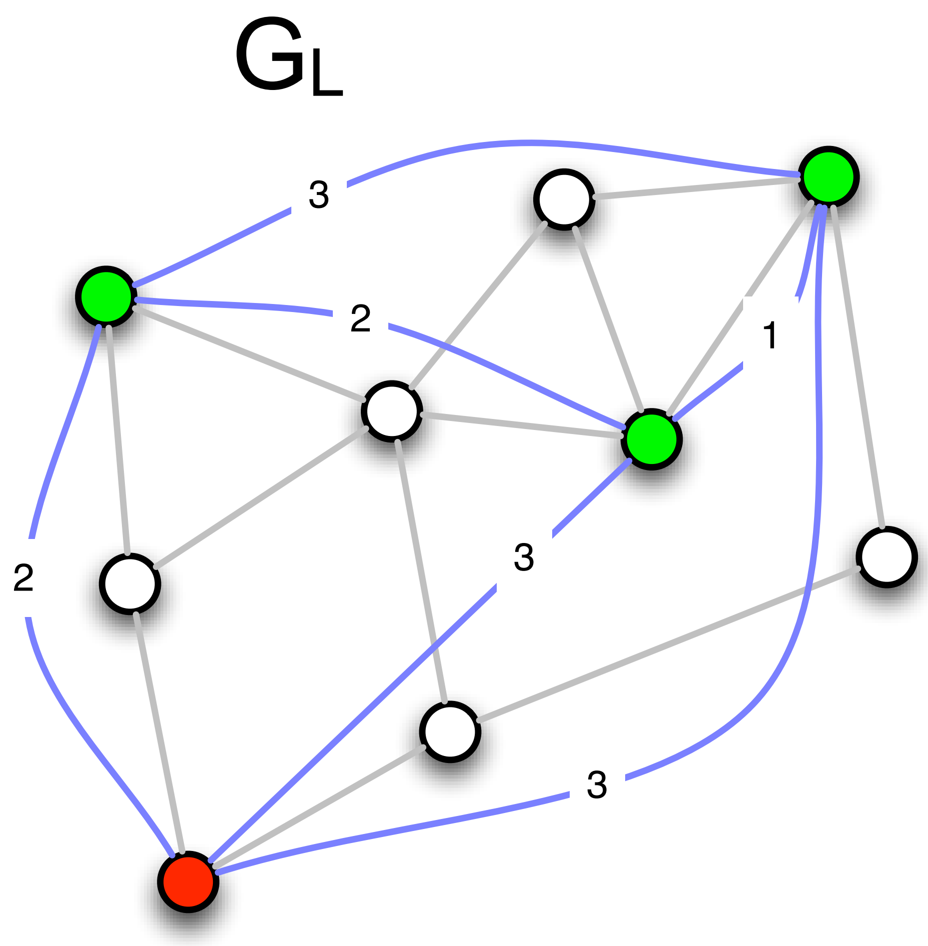
- For each edge **e=(u,v)** of the **MST**
  - Find shortest path P from u to v in G
  - If less than two nodes of P are in T then
    - Insert P into T
  - Else
    - Let p and q be the first and last node of P in T
    - Insert sub-path (u,p) and sub-path (q,v) of P into T
- **Output: Steiner tree approximation T**

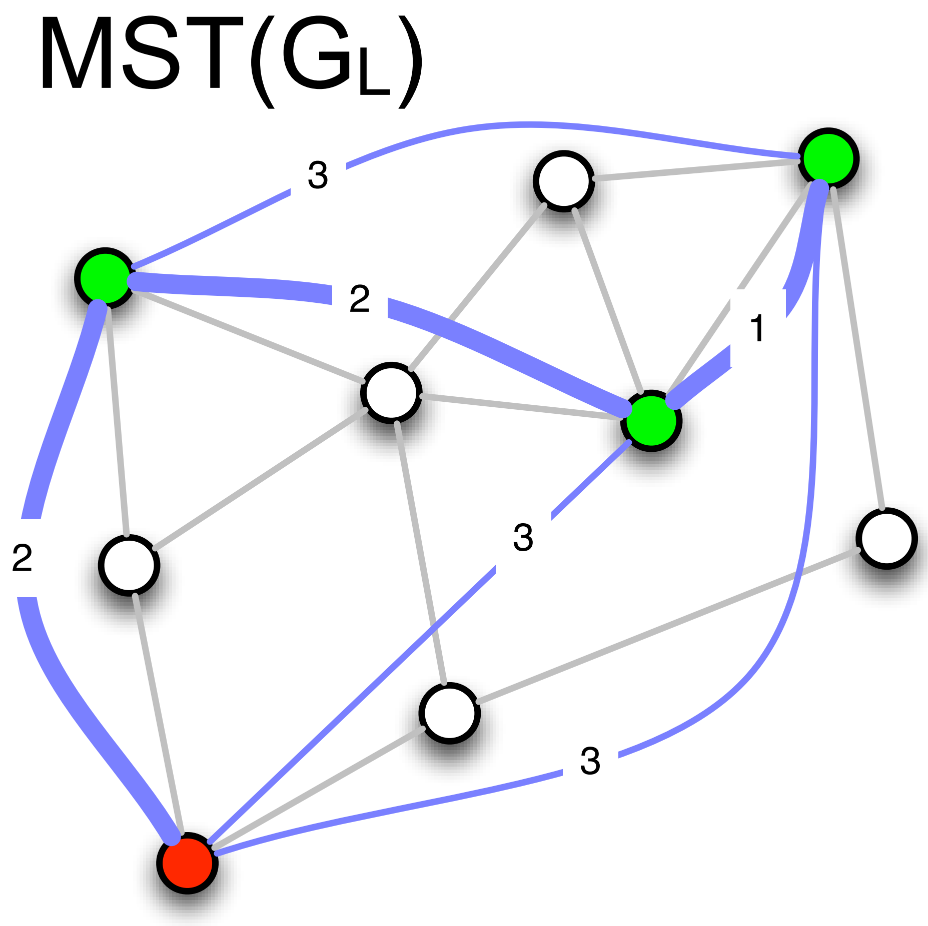
MST Steiner Tree Approximation Example





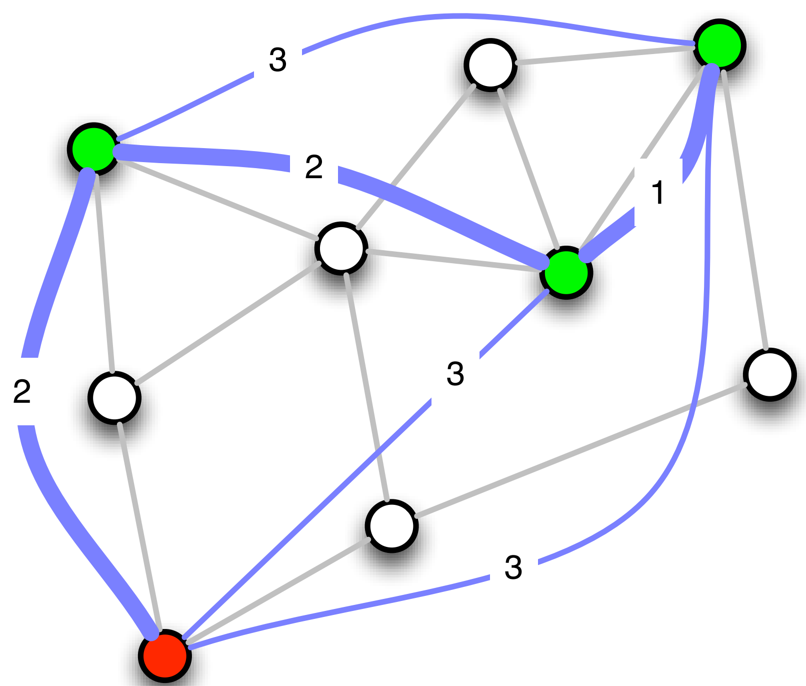
MST Steiner Tree Approximation
Example



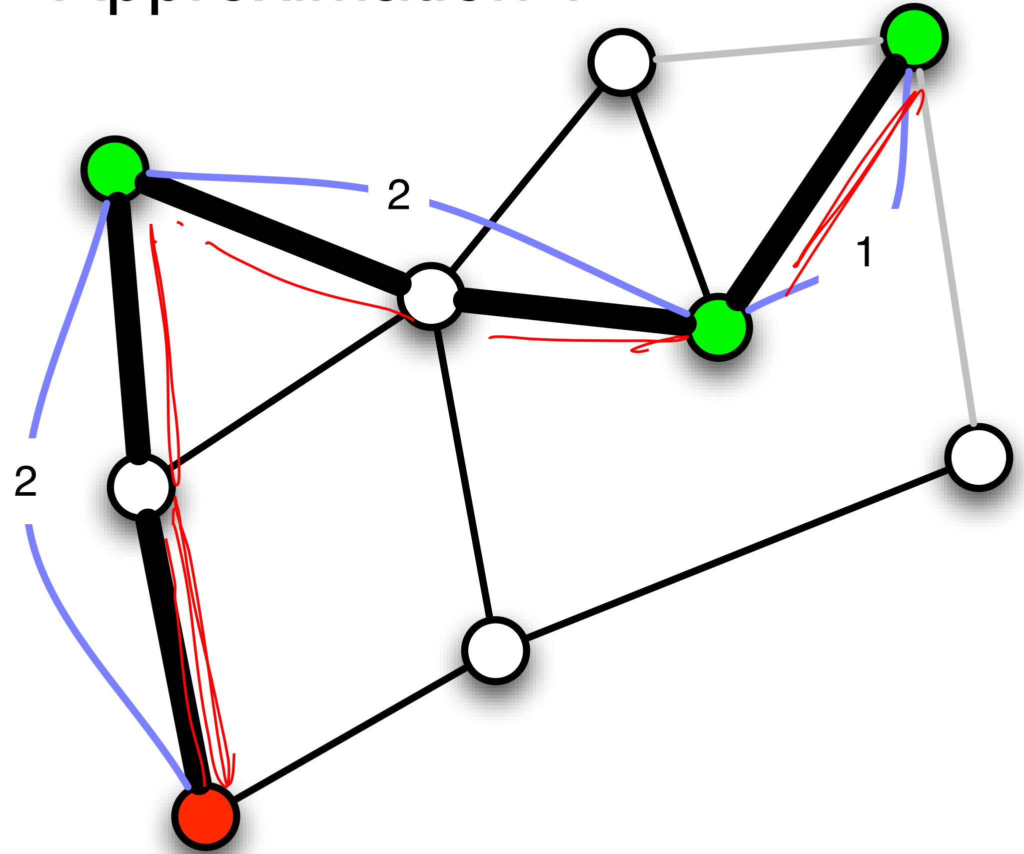


MST Steiner Tree Approximation Example

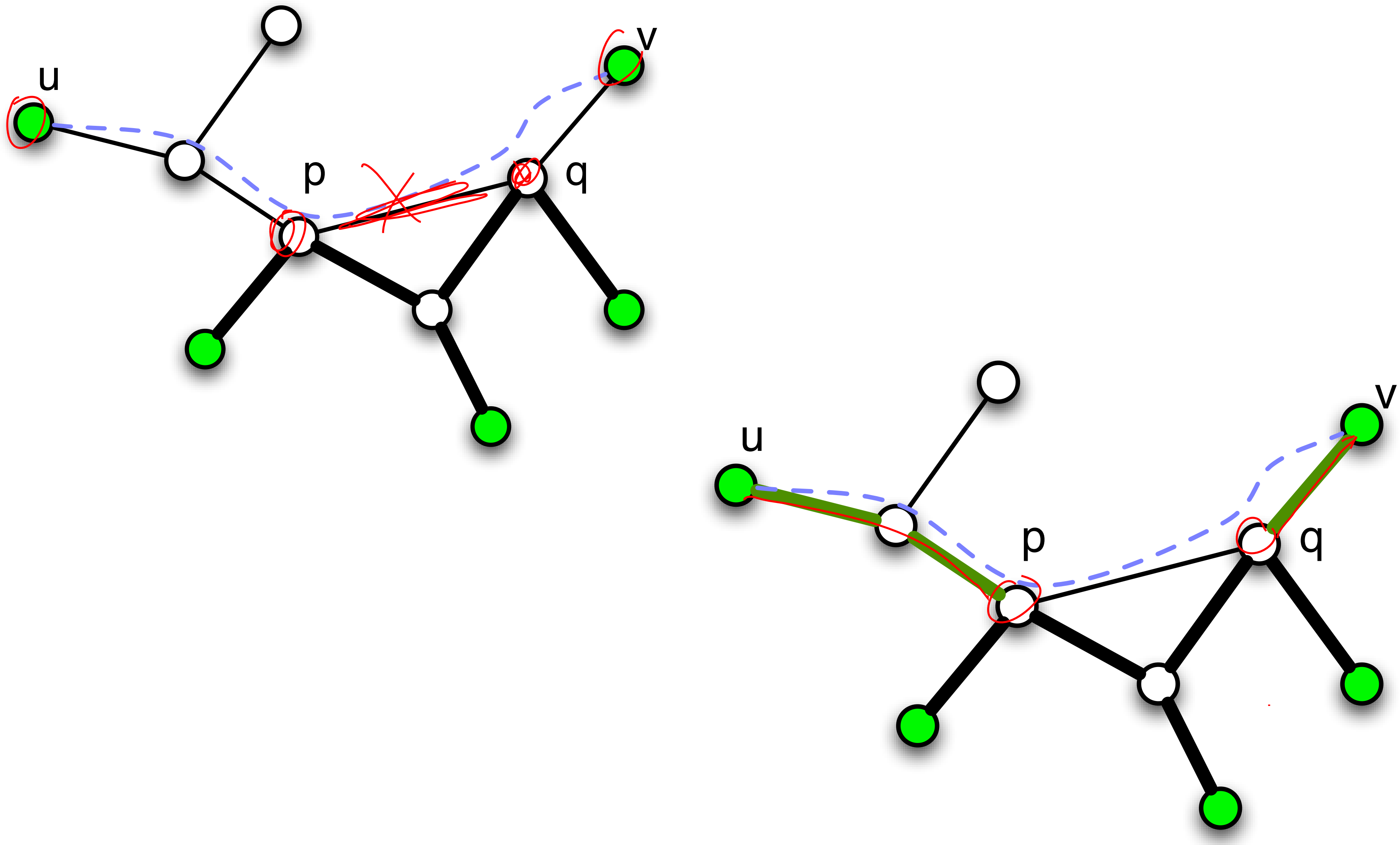
MST(G_L)



Steiner Tree Approximation T



Preventing Cycles



Quality of the MST-Approximation

- **Theorem**
  - The MST-Approximation constructs a tree in polynomial time. The edge sum of this tree is at most twice as large as those of the Steiner tree.
- **Proof sketch**
  - A Steiner Tree without Steiner points is a factor 2 approximation of the Steiner tree
  - The minimum spanning tree of G_L is the Steiner Tree without Steiner points of G
  - This results in approximation algorithm with factor 2

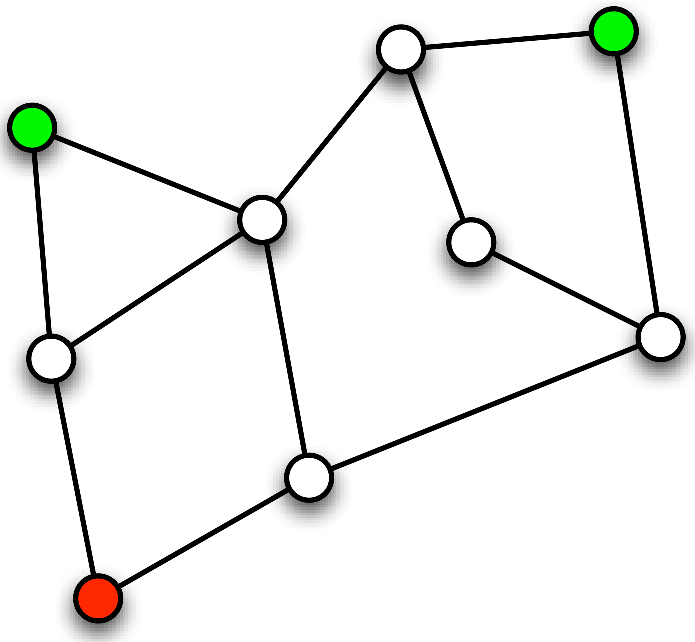
Routing Models for Data Aggregation

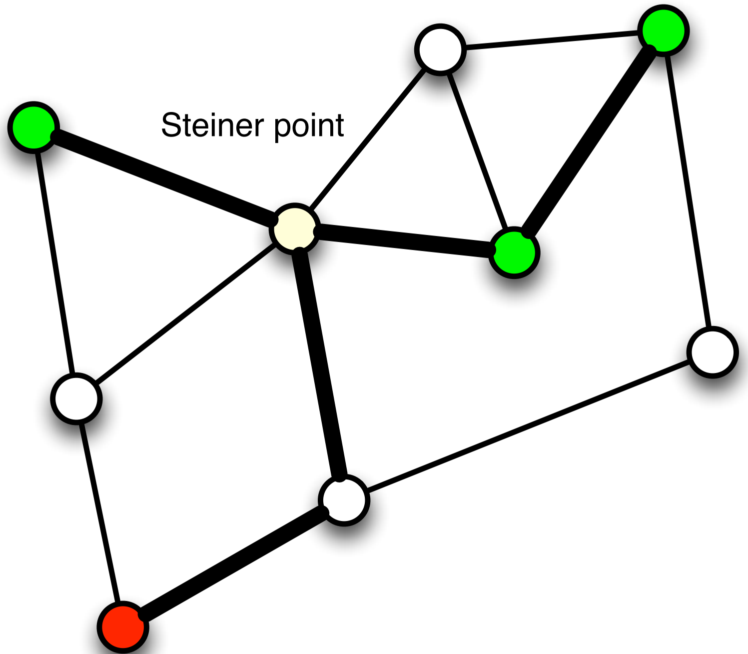
- **Address Centric Protocol**
  - each sensor sends independently towards the sink
  - not suitable for (real) aggregation
- **Data Centric Protocol**
  - Forwarding nodes can read and change messages
- **Literature**
  - Krishnamachari, Estrin, Wicker The Impact of Data Aggregation in Wireless Sensor Networks, Proc. of the 2nd Int. Conf. on Distributed Computing Systems Workshops (ICDCSW'02)

Energy Optimal Tree Structure

- **Given:**
  - set of data sources and a sink
  - communication graph G
- **Compute:**
  - Steiner tree T
    - sub-graph of G
    - connects all sources and sinks
    - number of edges is minimal
- **Alternative:**
  - edges have an (energy) weight
  - minimize the sum of edge weights

Steiner Tree Problem





Theoretical Bounds

- **Costs for address based Routing N_A**

$$N_A = \sum_i d_i$$

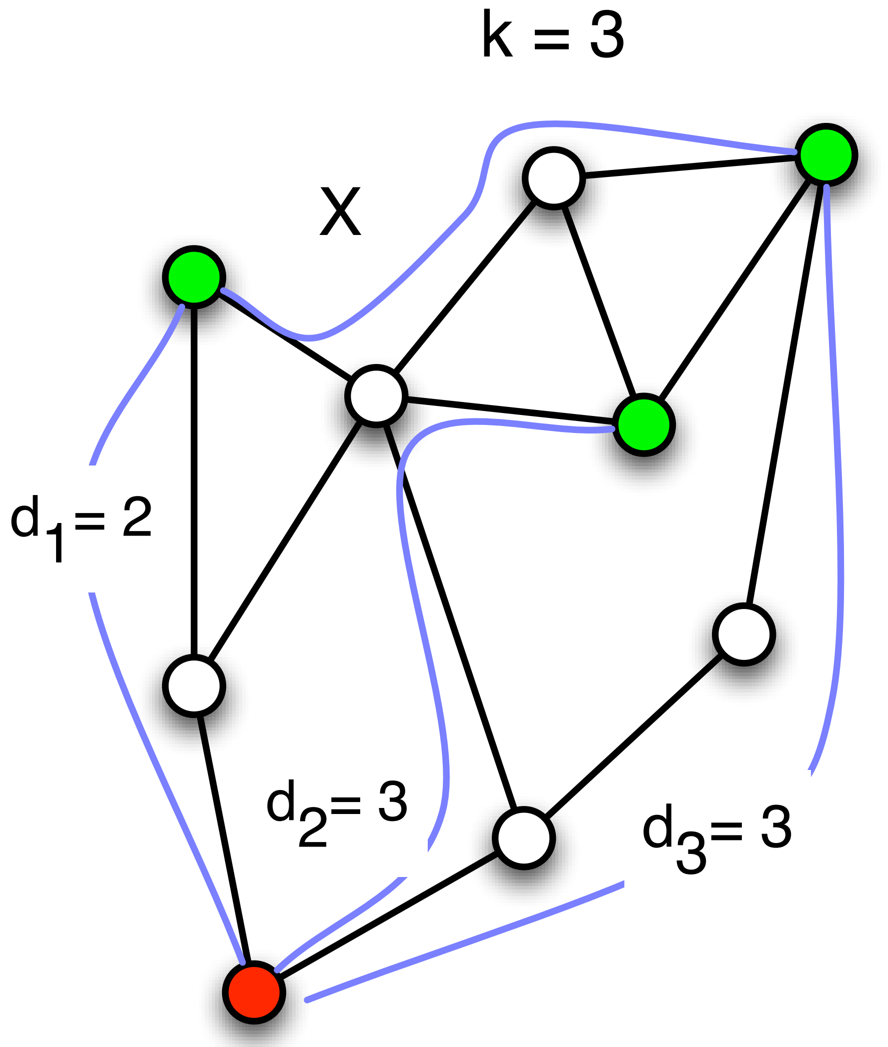
  - d_i: shortest distance form source i to sink s

- **Cost for optimal data centric routing N_D = weight of Steiner-tree**

$$N_D \leq (k-1)X + \min_i\{d_i\}$$

  - X: maximal shortest path between sources
  - k: number of sources

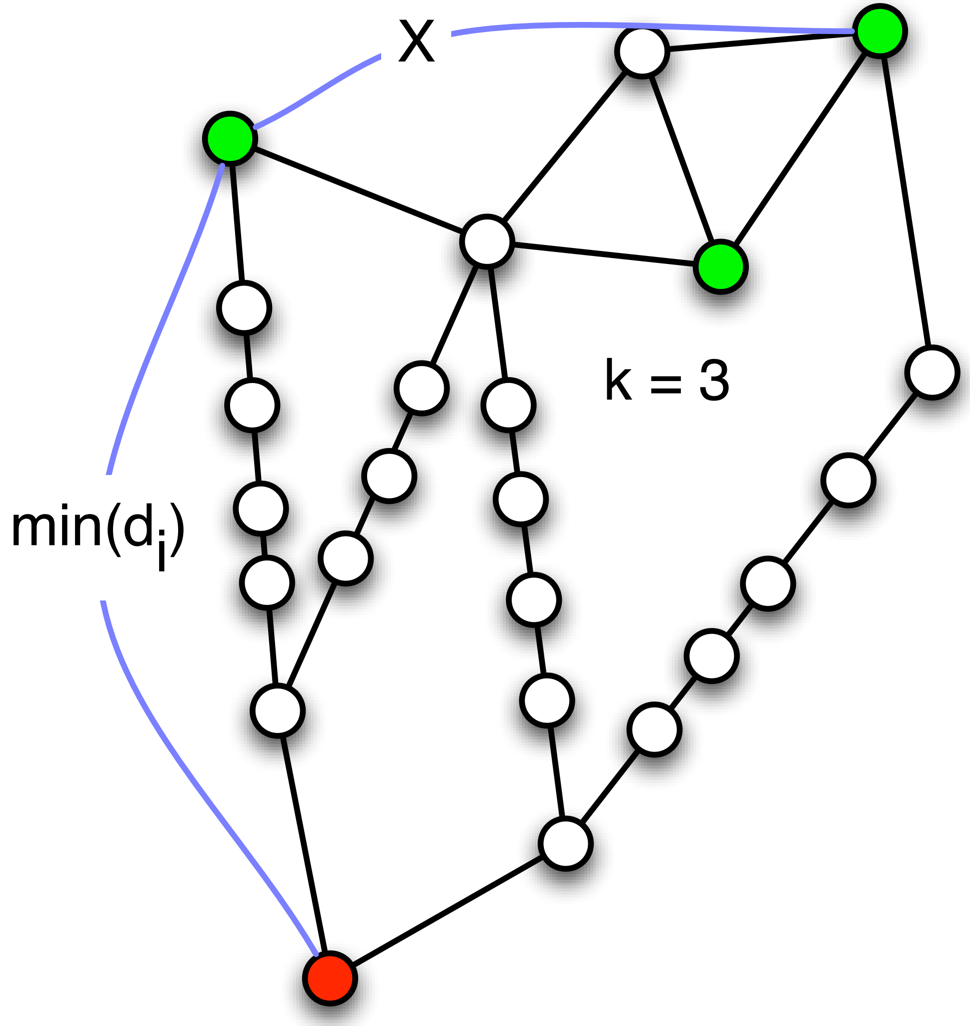
$$N_D \geq \min_i\{d_i\} + k - 1$$



Theoretical Bounds

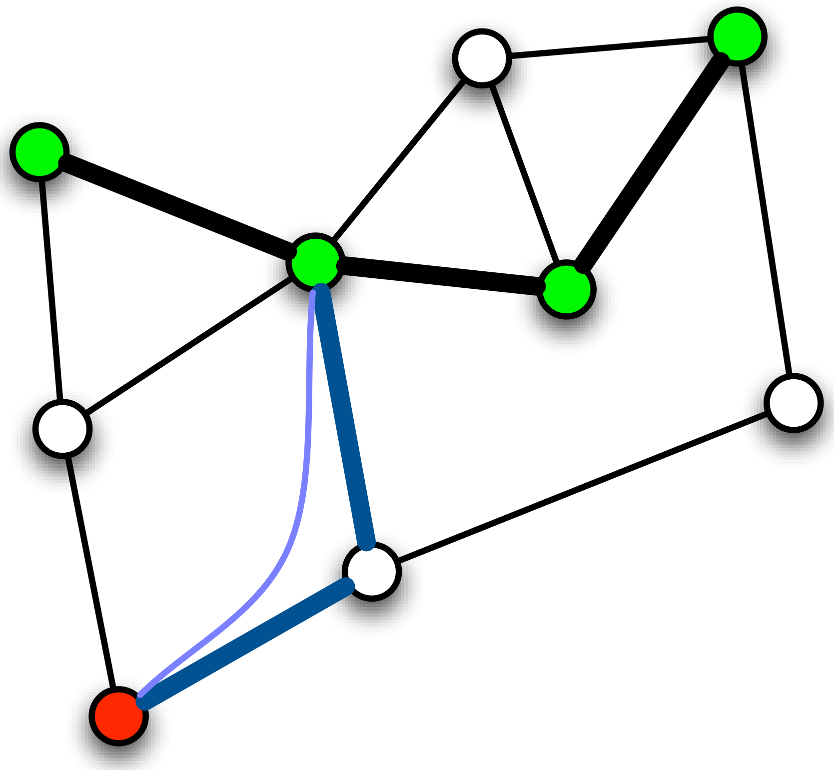
- **For fixed X and k and growing $\min_i\{d_i\}$**

$$\lim_{d \to \infty} \frac{N_D}{N_A} = \frac{1}{k}$$



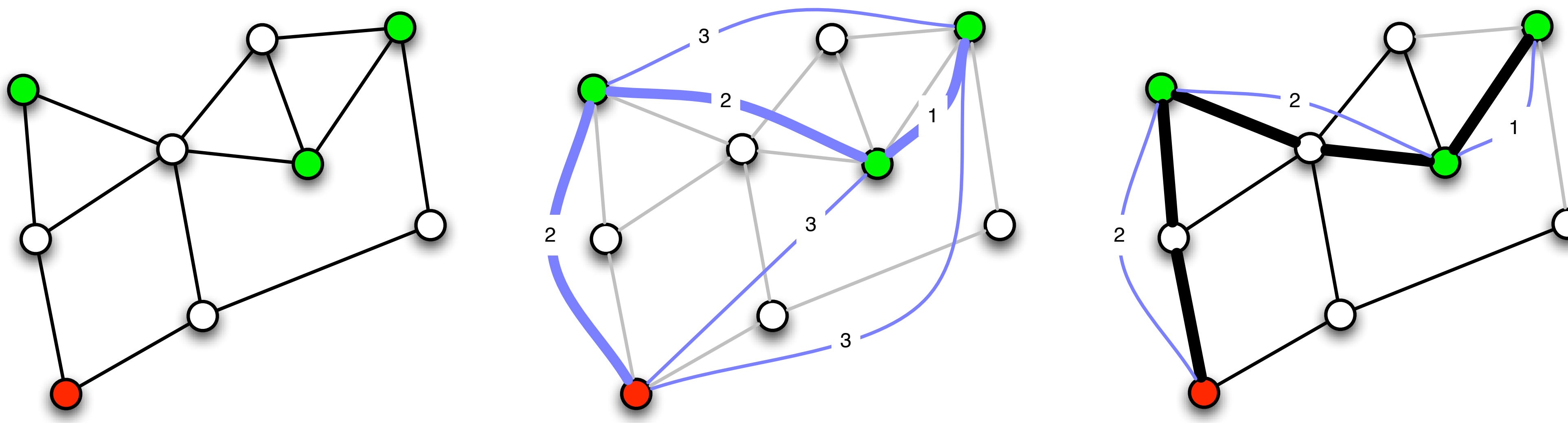
Theoretical Bounds

- **Theorem**
  - If the subgraph induced by the sources is connected, then the optimal routing can be computed in polynomial time
- **Proof sketch**
  - Compute MST T for the sources
  - Compute the shortest path from T to the sink



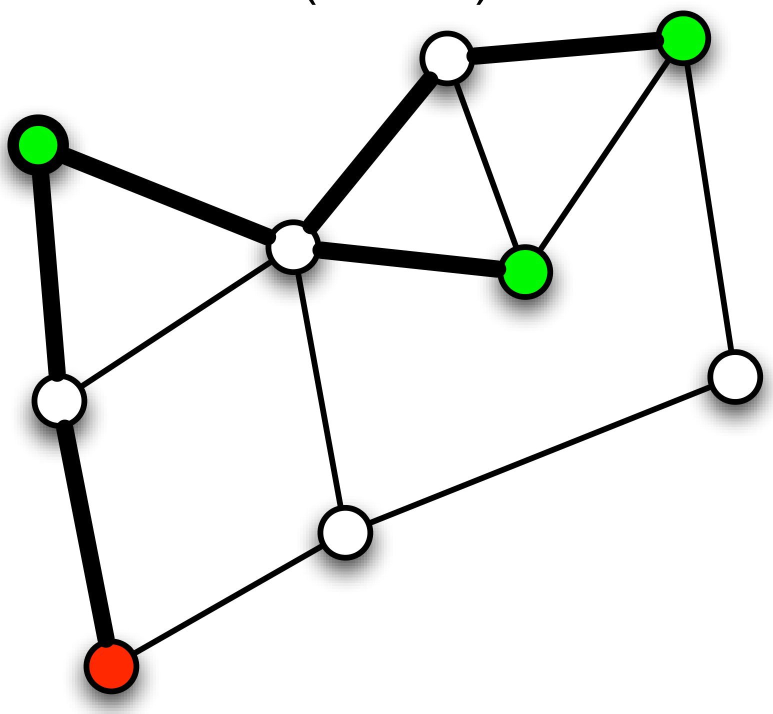
Approximation Algorithm

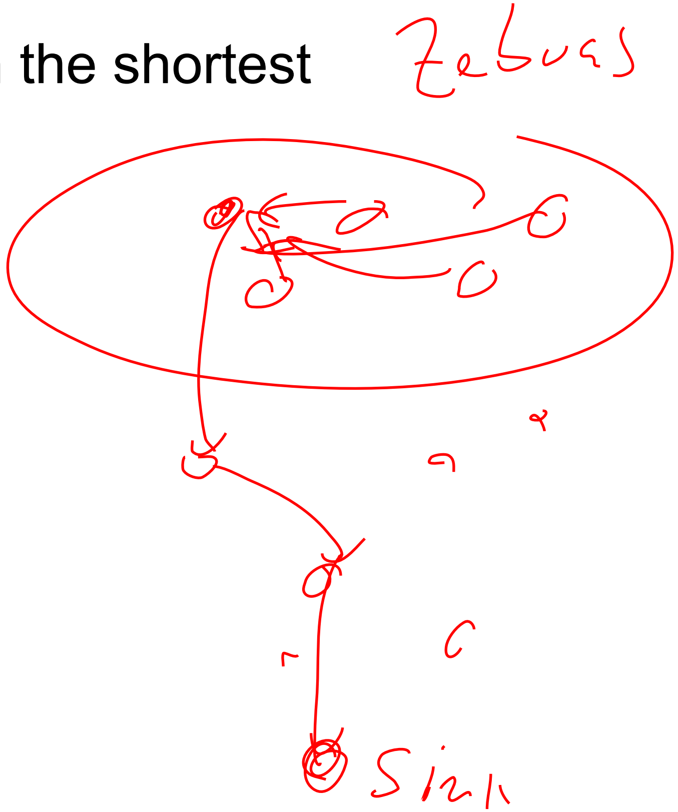
- **The Steiner tree approximation algorithm (of the last lecture) cannot be implemented efficiently in a WSN**



Suboptimal Aggregation

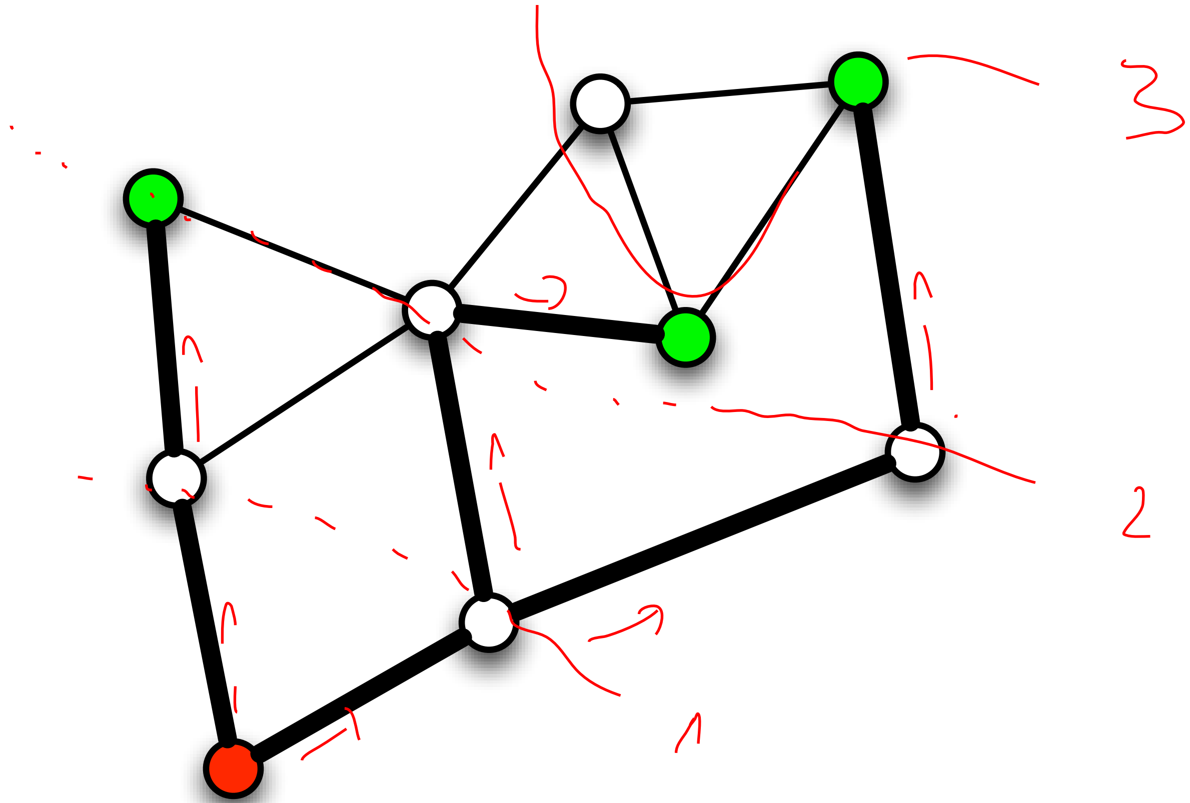
- **Center at Nearest Source (CNS)**
  - Data source closest to the sink collects all information
  - All other sources send the information on the shortest path to this source (center)





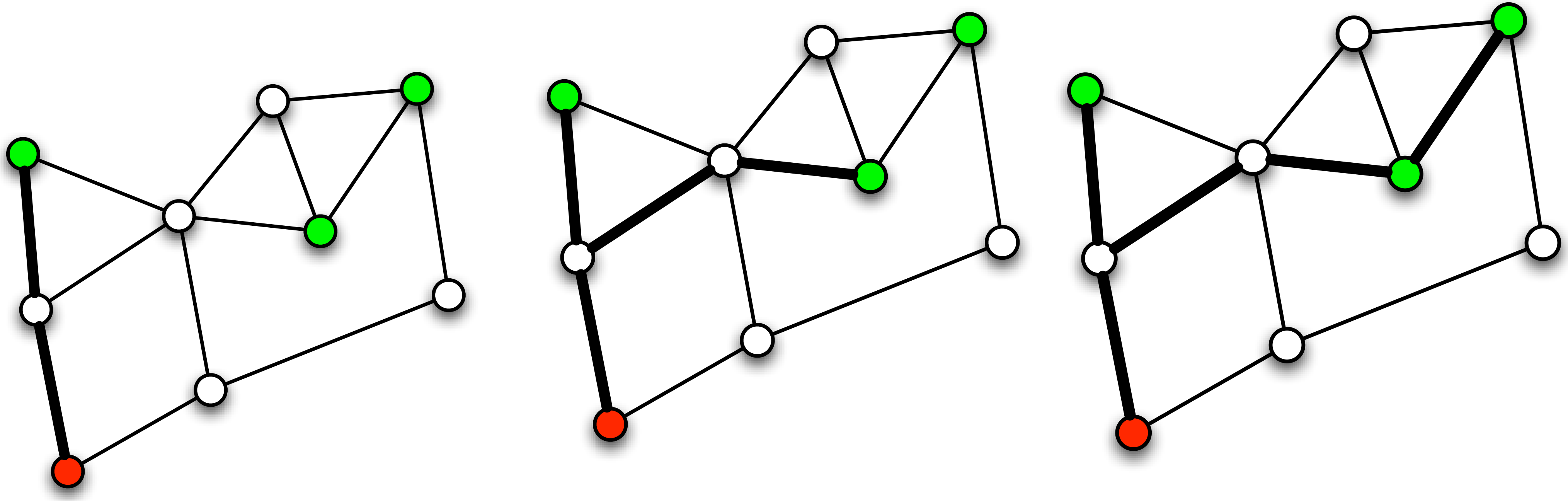
Suboptimal Aggregation

- **Shortest Paths Trees (SPT)**
  - Set of all shortest paths from the sources to the sink

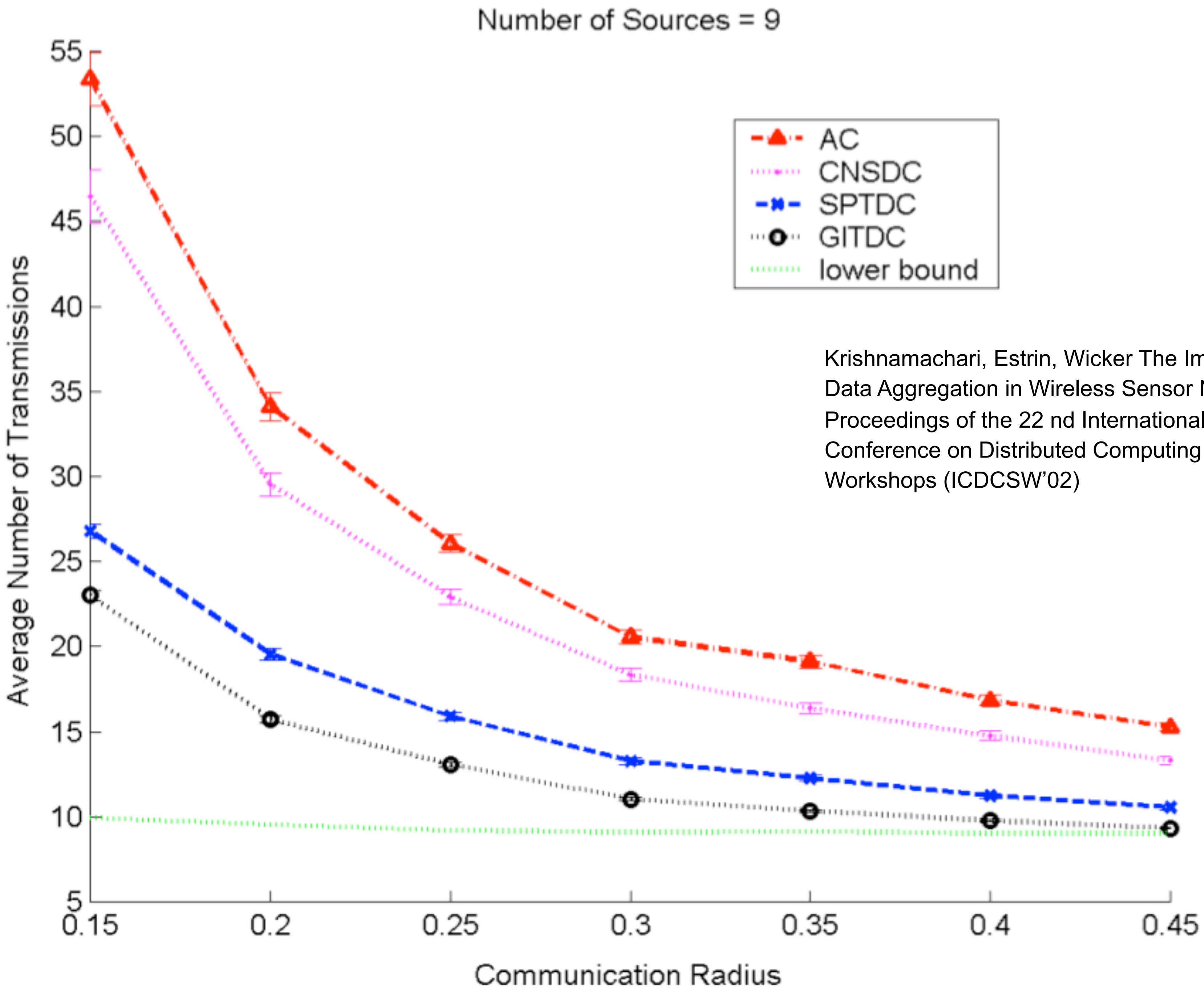


Suboptimal Aggregation

- **Greedy Incremental Tree (GIT)**
  - Select the shortest path between the data source, closest to the sink, and the sink
  - Select successively the closest node to the tree and the shortest path to any of the tree nodes

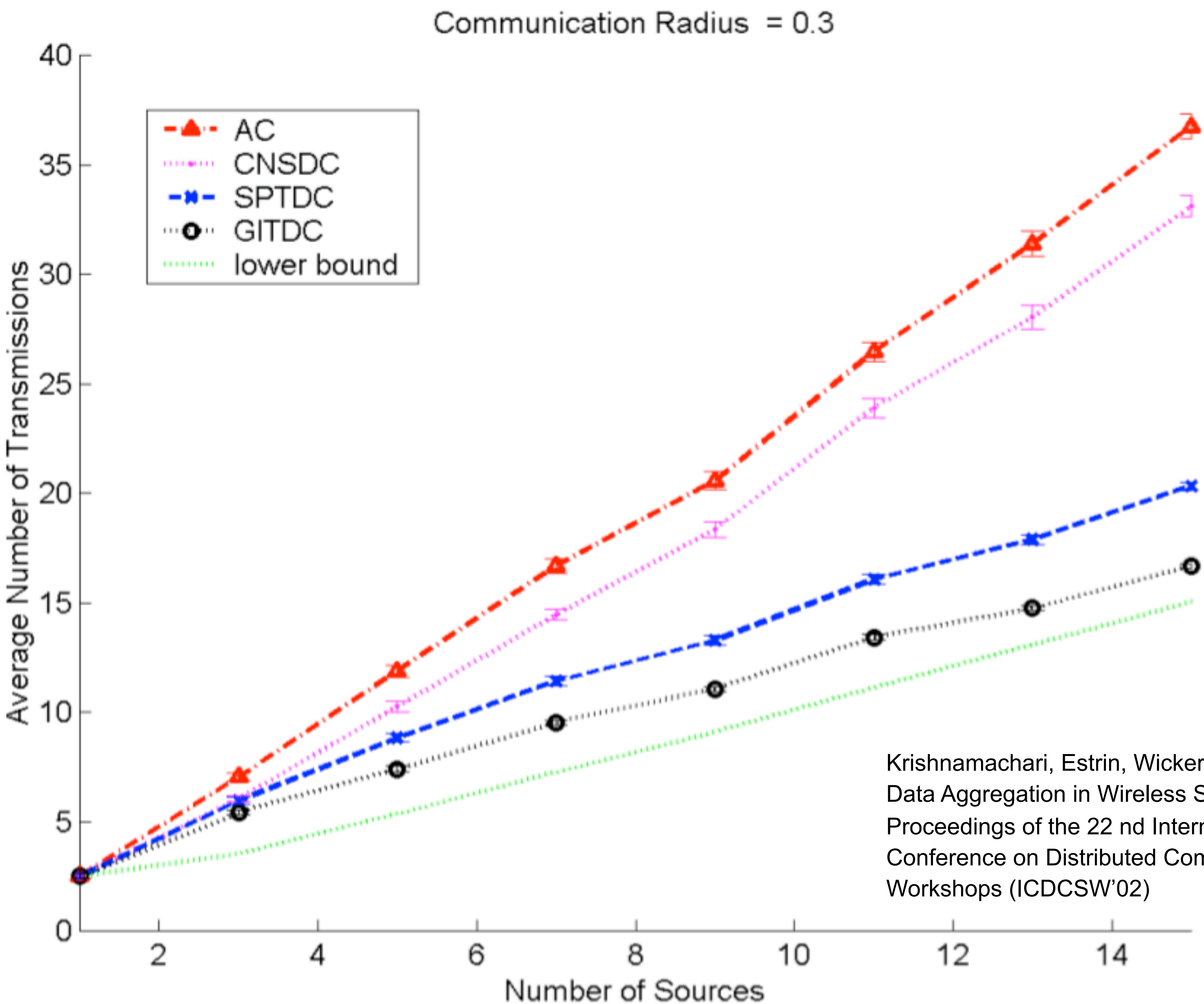


Energy Saving by Data Aggregation



Krishnamachari, Estrin, Wicker The Impact of Data Aggregation in Wireless Sensor Networks, Proceedings of the 22 nd International Conference on Distributed Computing Systems Workshops (ICDCSW'02)

Energy Saving by Data Aggregation



Krishnamachari, Estrin, Wicker The Impact of Data Aggregation in Wireless Sensor Networks, Proceedings of the 22 nd International Conference on Distributed Computing Systems Workshops (ICDCSW'02)

ALBERT-LUDWIGS-
UNIVERSITÄT FREIBURG

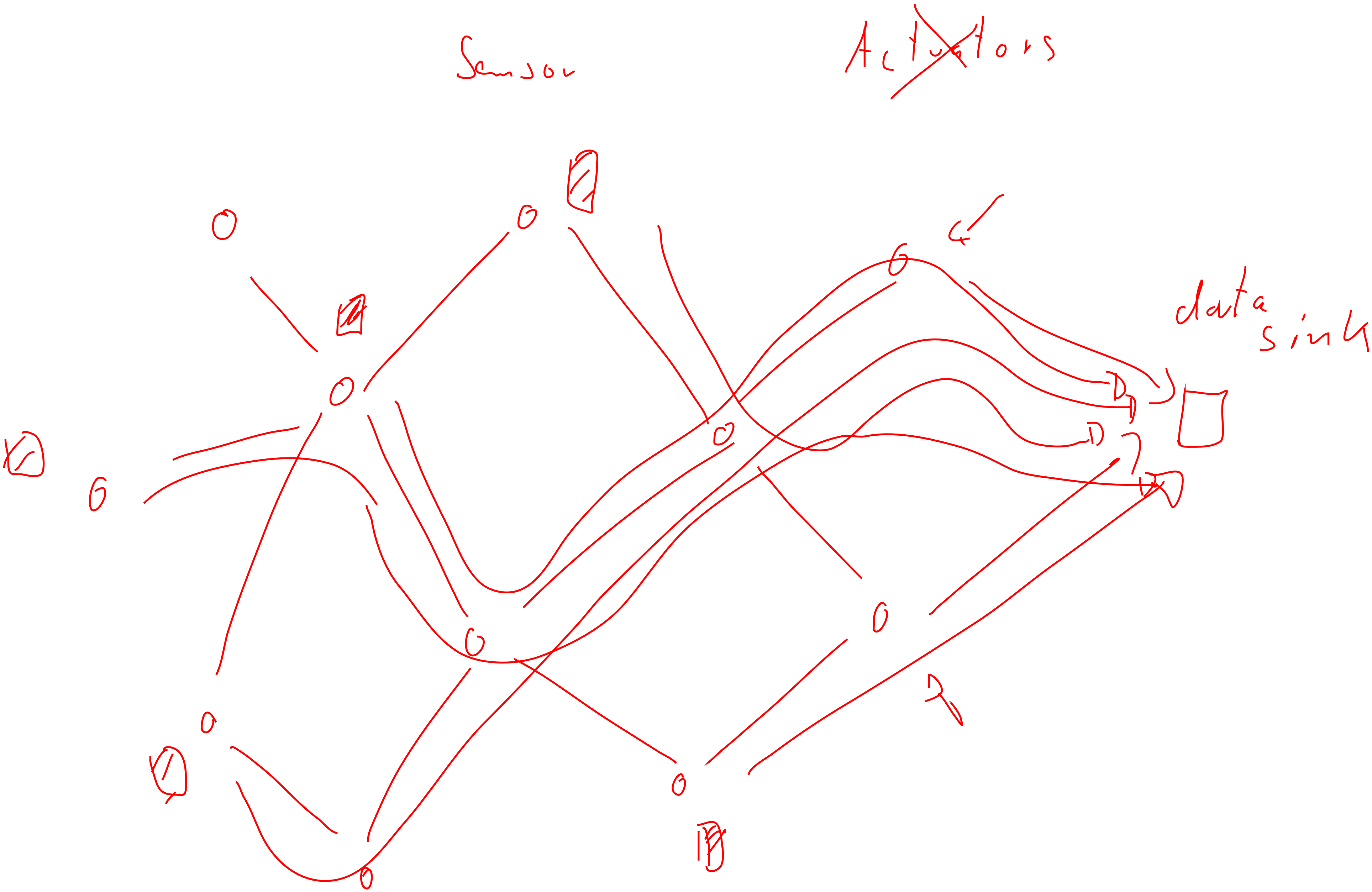
Algorithms for Radio Networks

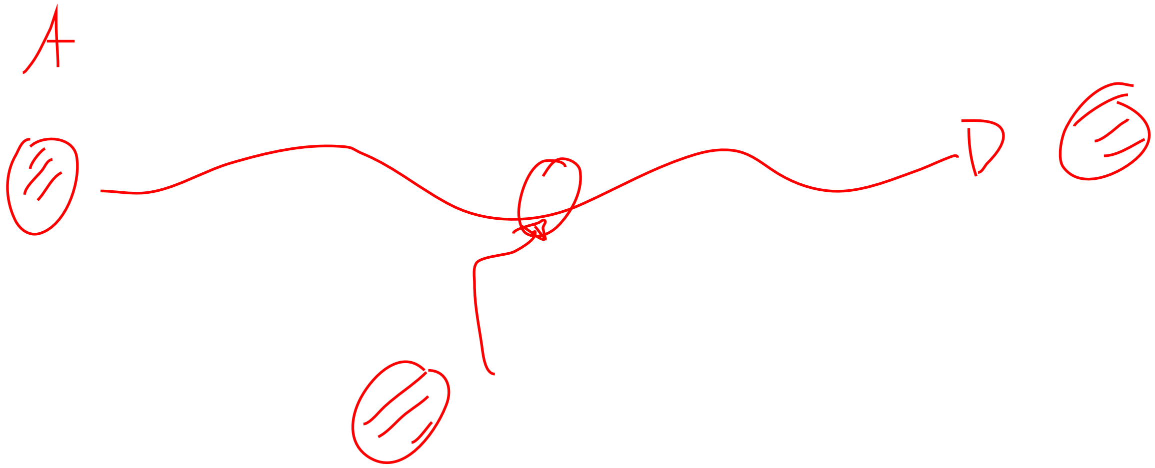
Data Aggregation

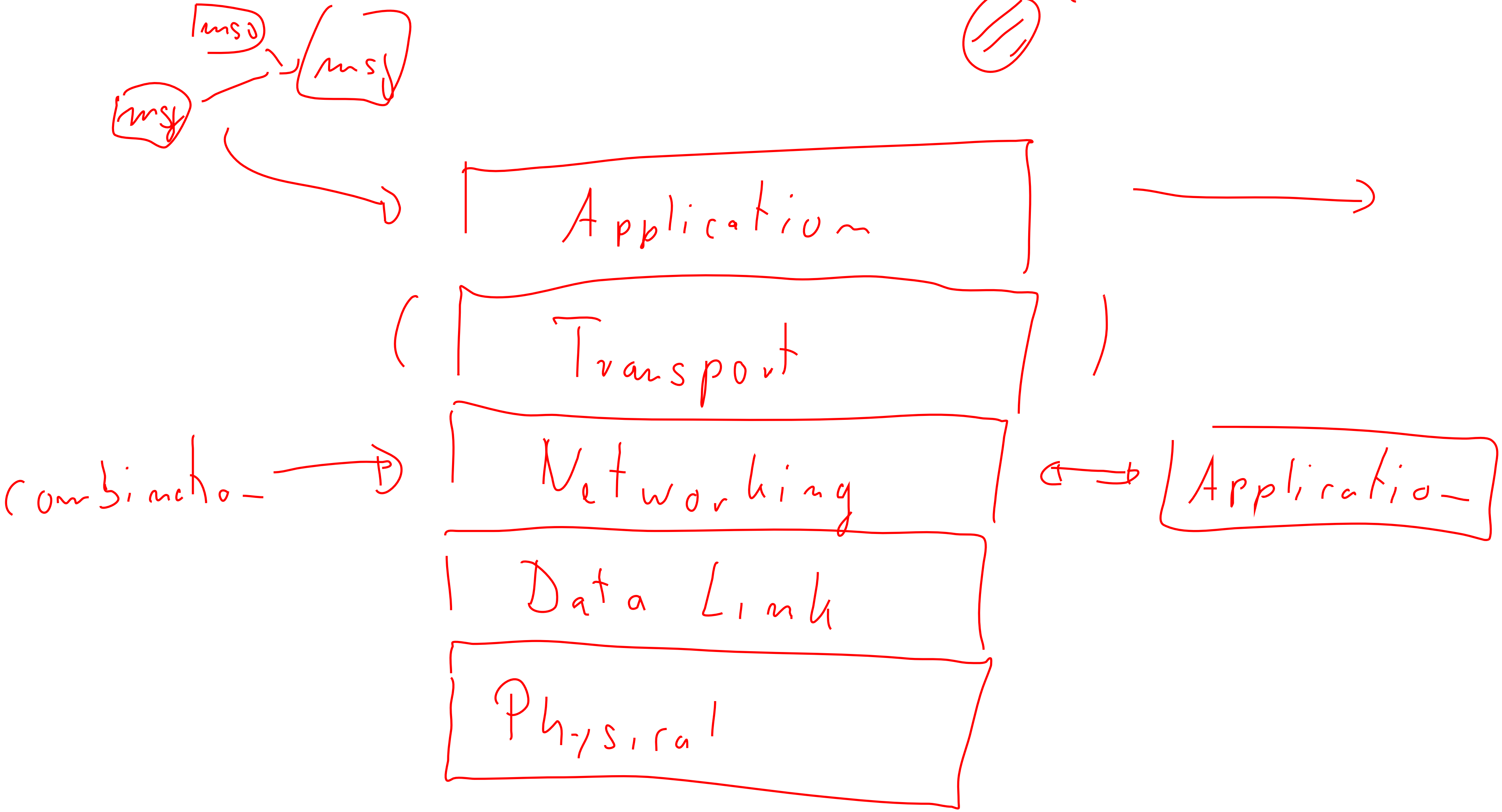
University of Freiburg
Technical Faculty
Computer Networks and Telematics
Christian Schindelhauer

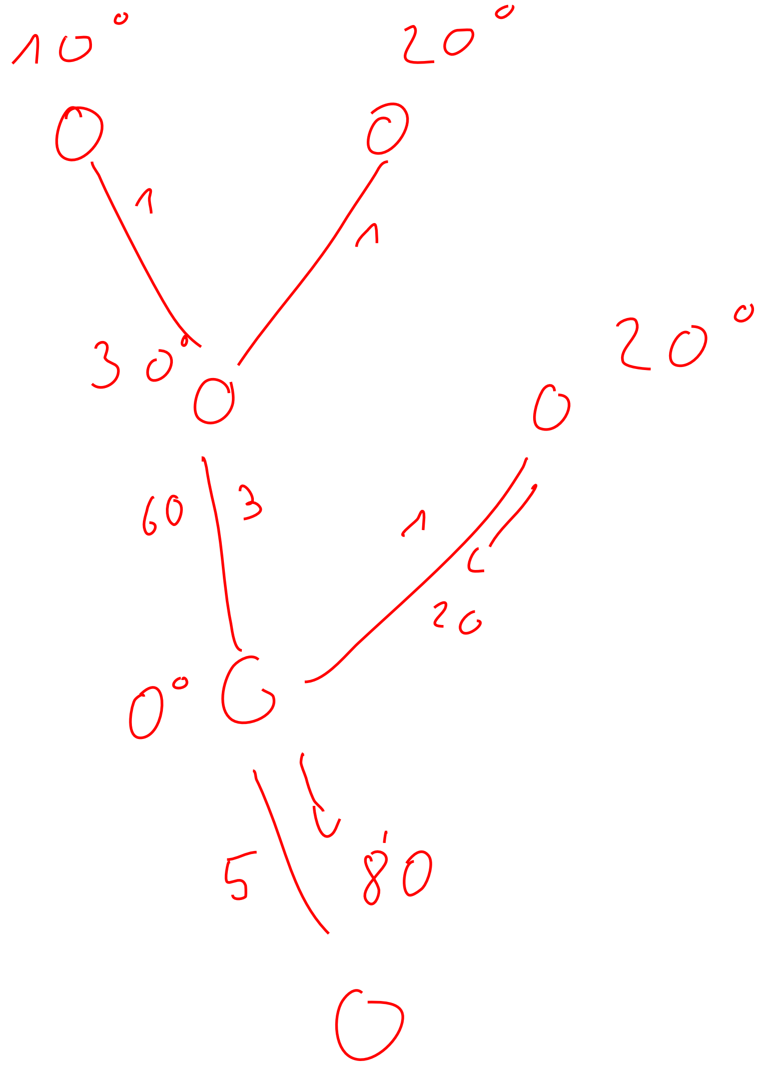
CoNe Freiburg

IIF
INSTITUT FÜR
INFORMATIK
FREIBURG


Sensor
~~Actuators~~
data sink


A
D


msg
msg
msg
Application
Transport
Networking
Data Link
Physical
combination
Application



Datasing

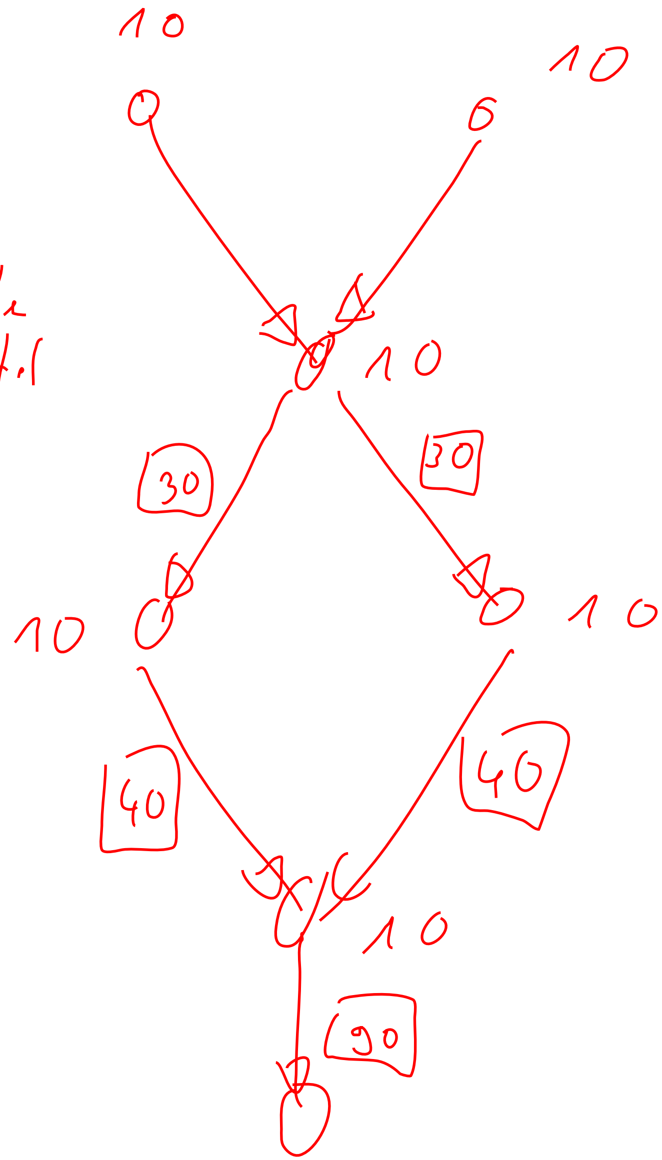
$$\frac{10+20+30+20+0}{5} = \frac{80}{5} = \underline{16°}$$

connected

DAG

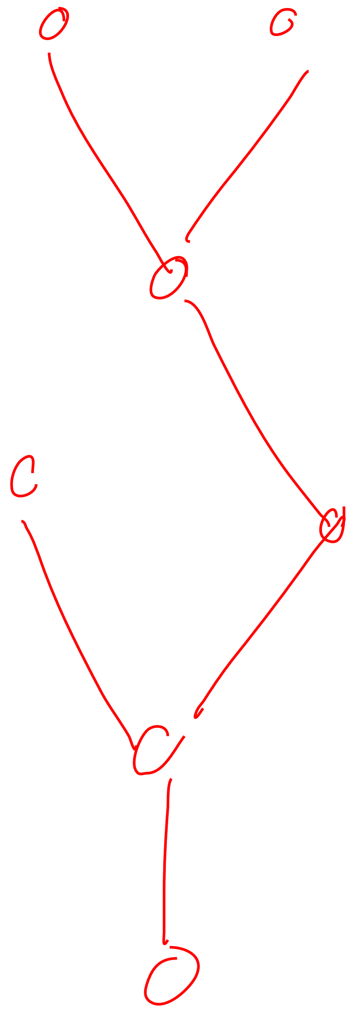
without cycle in the undirected version

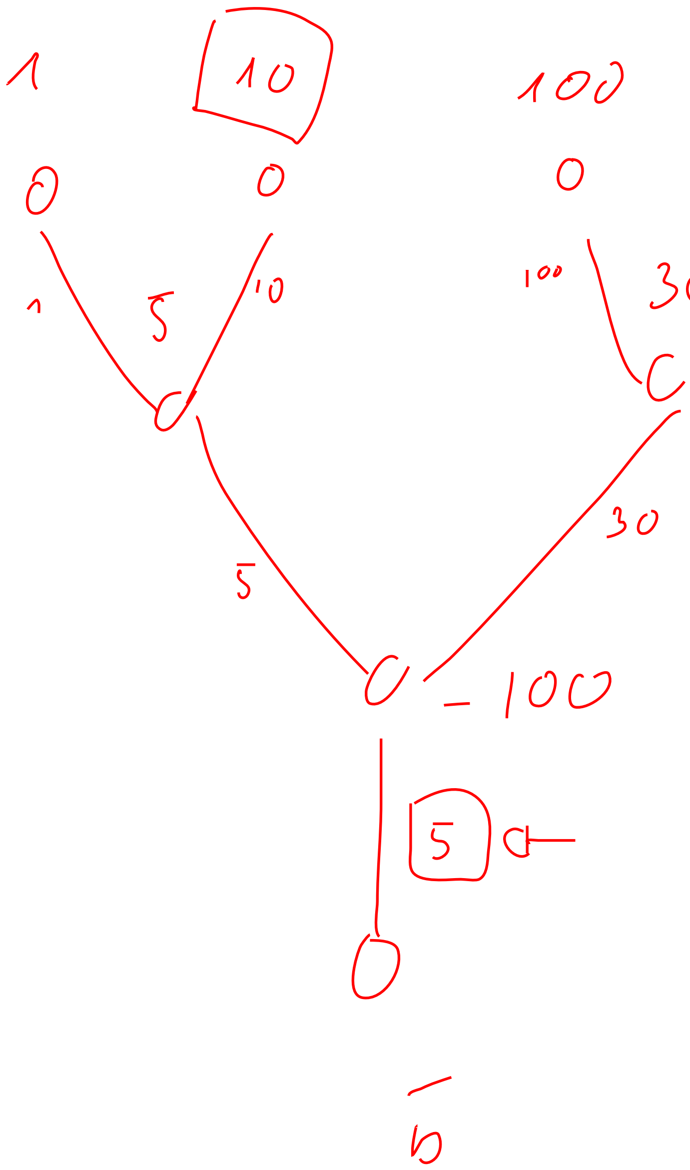
TREE



60

90





data < 5

bound the # of msgs

median

25% 50% quantile 75% - quantile

complicated

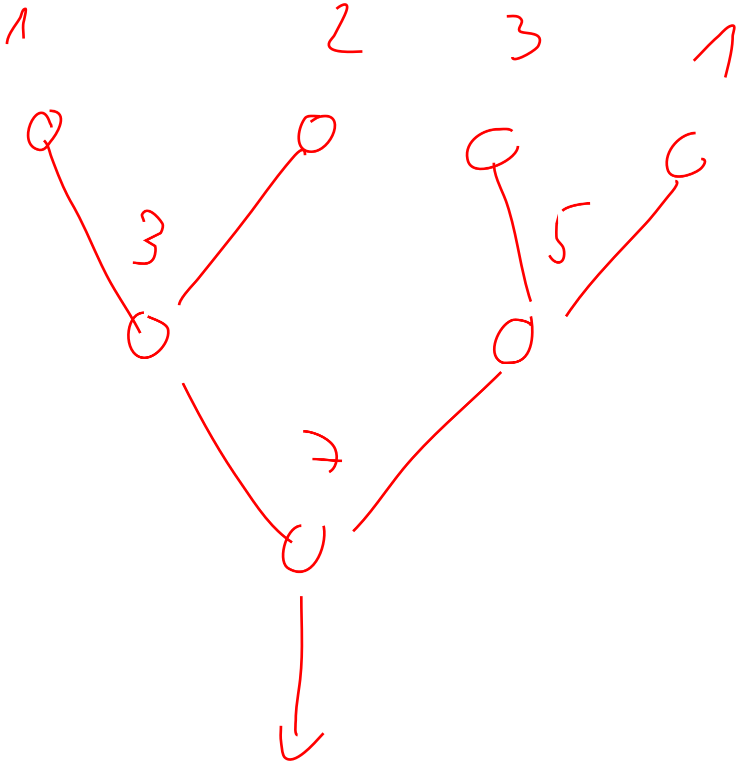
easy easy

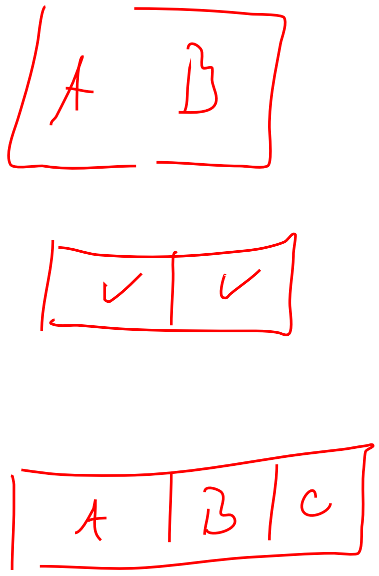
1 10 20 30 100 110 130

95% quantile

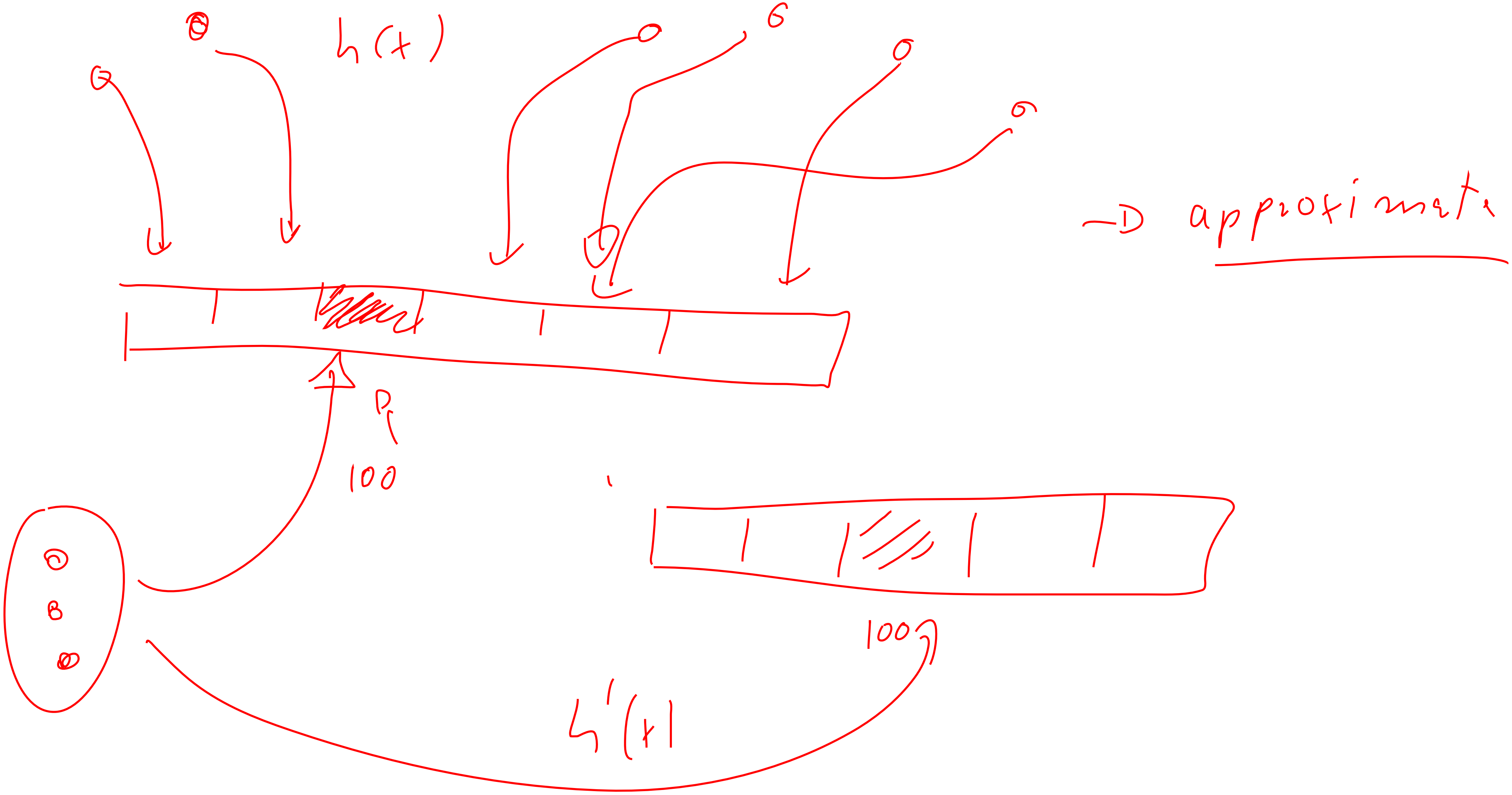
5% quantile

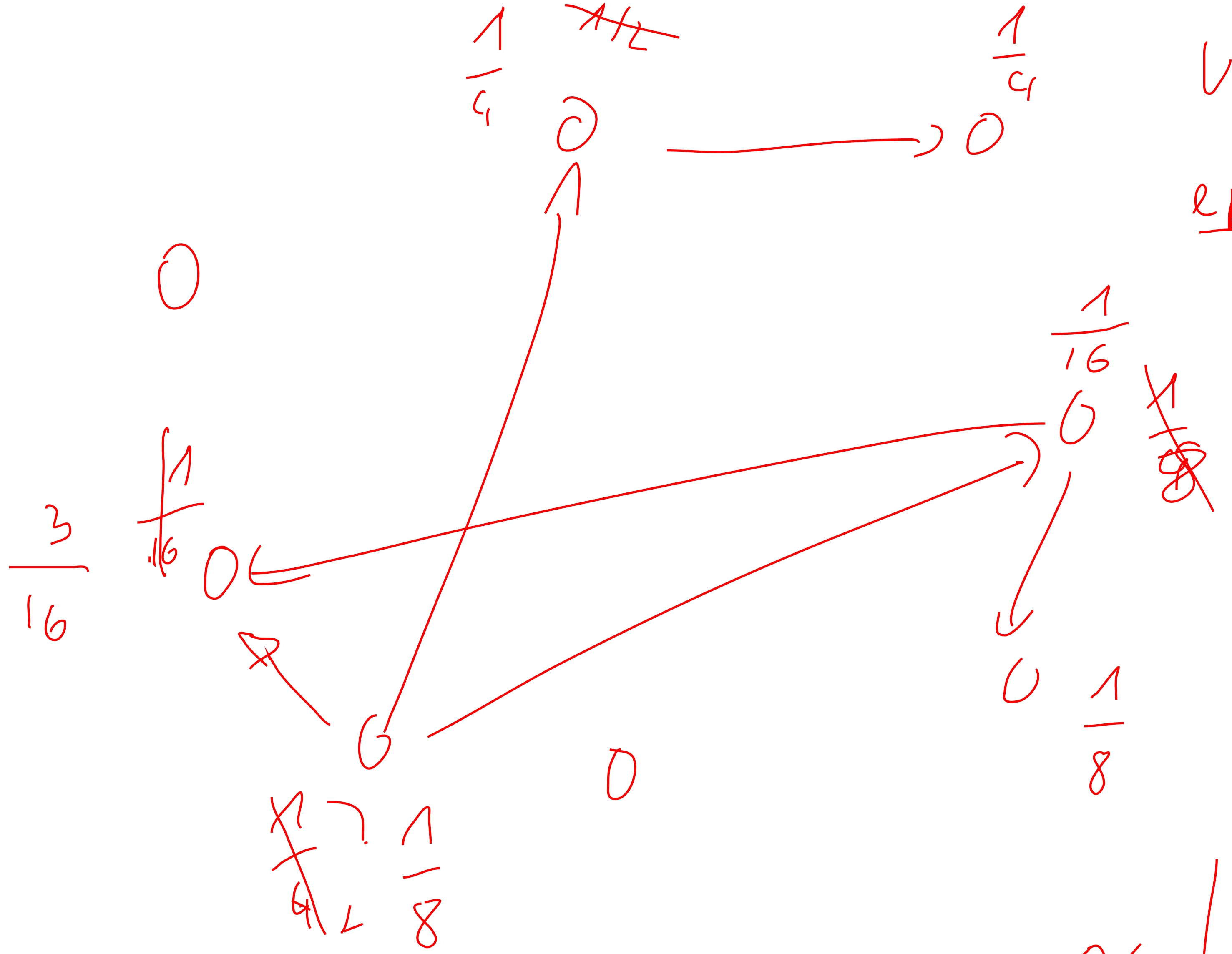
$\frac{1}{20}$





5 different values


h(x)
→ approximate
100
100
h'(x)

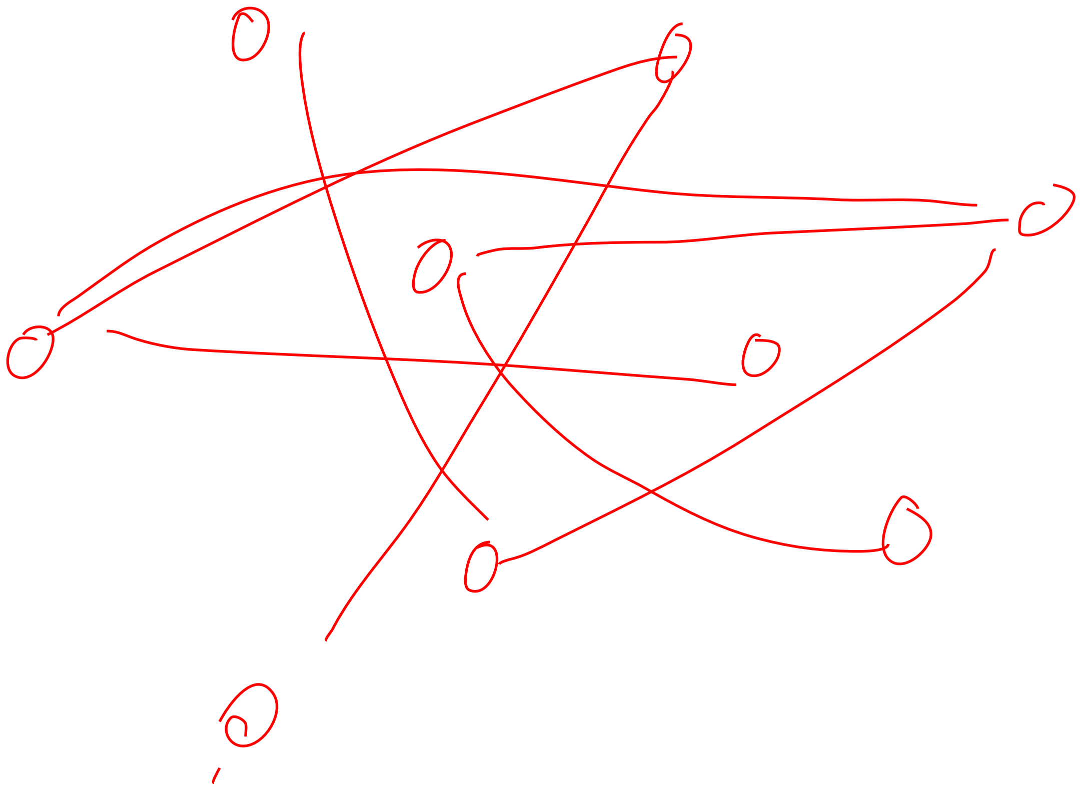


Virus

epidemic algorithms

$\rightsquigarrow \boxed{\frac{1}{n}}$

$\frac{1}{\frac{1}{n}} = n$

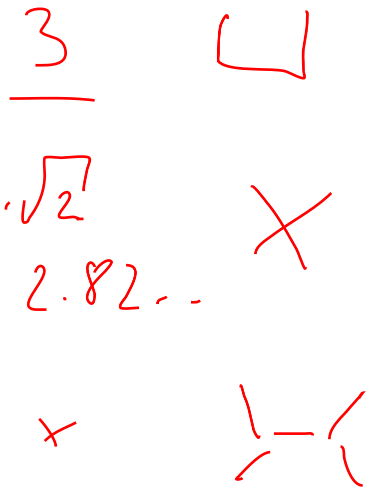


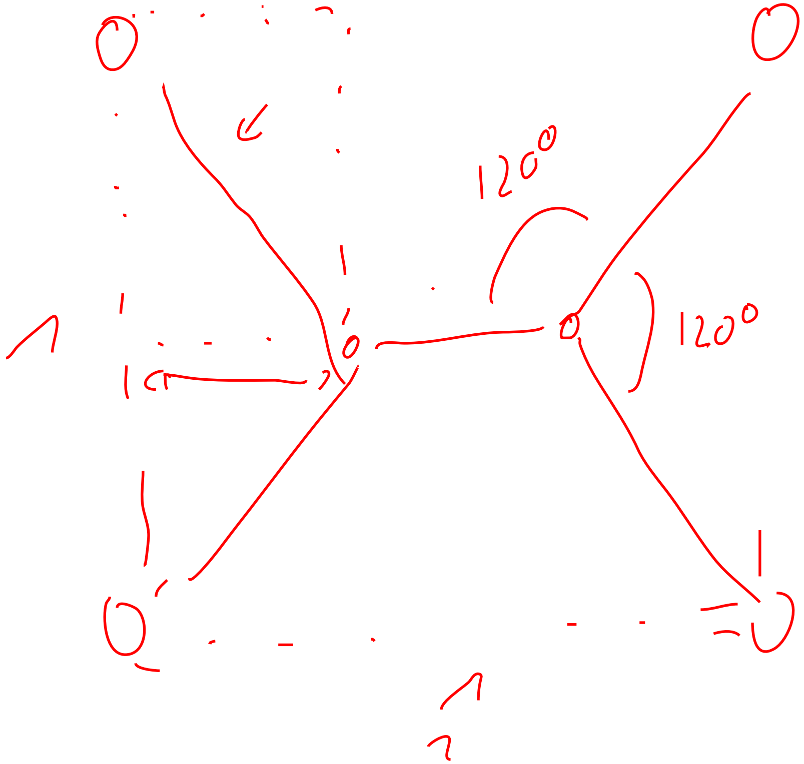

Terminal
Terminal

cost: $\underline{3}$

$2 \cdot \sqrt{2}$

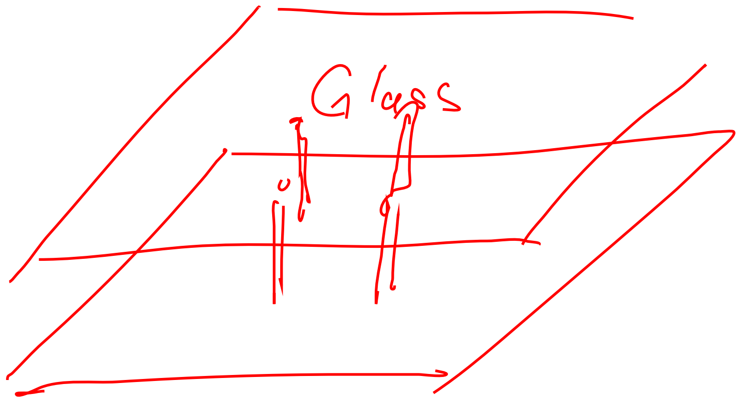
$\approx 2.82\ldots$



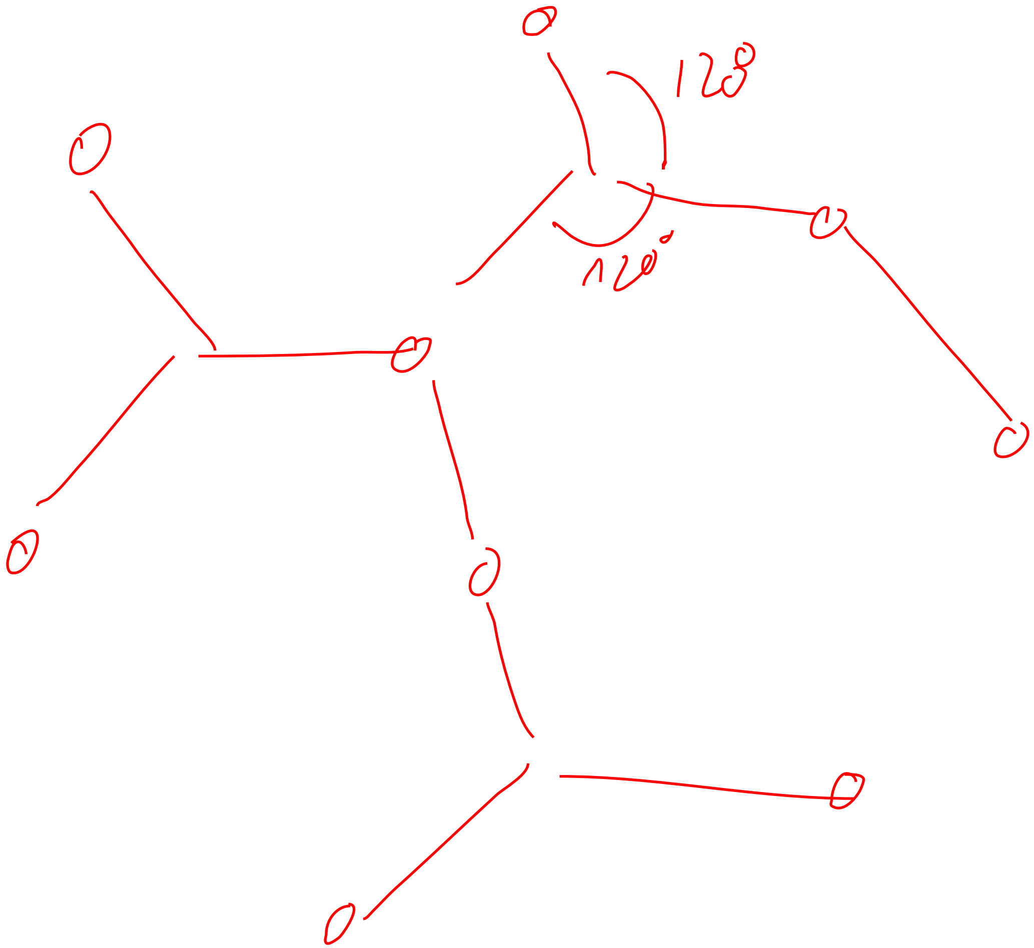


$\sqrt{1^2 + 1^2}$

$= 2\sqrt{2}$



Steiner Tree



NP-complete

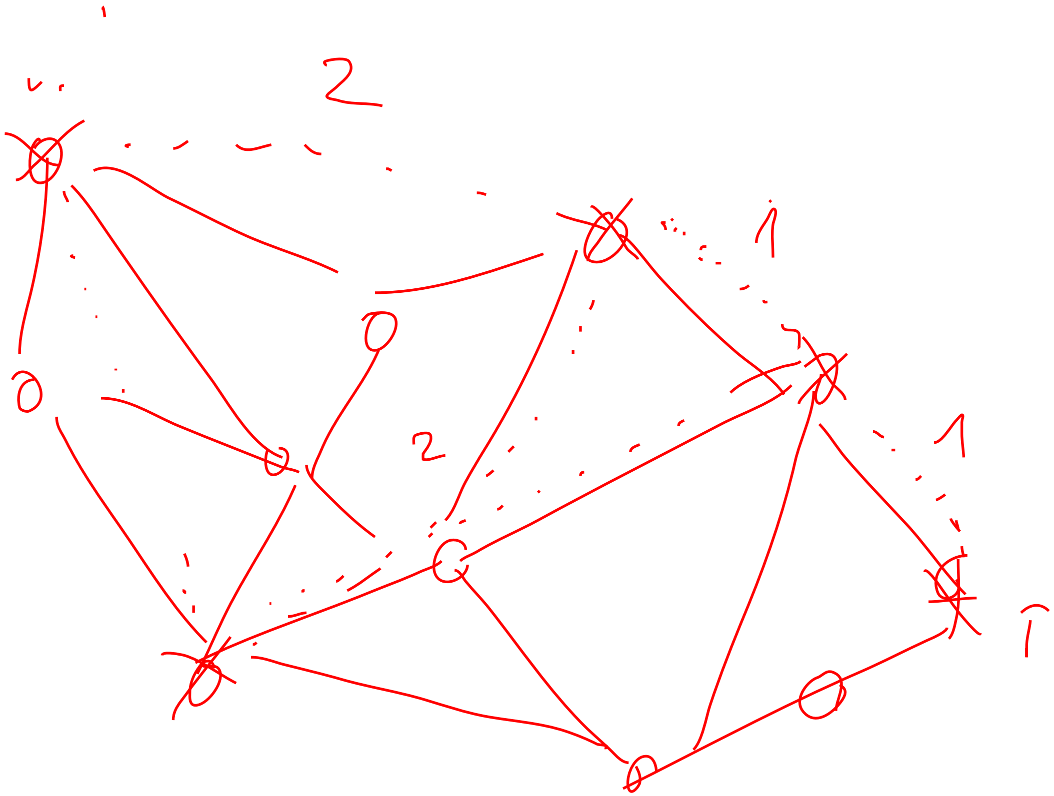
approximat.

up to a factor

of $1+\varepsilon$

in time $n^{c\left(\frac{1}{\varepsilon}\right)^{c}}$

MST



- Tree
- Spanning

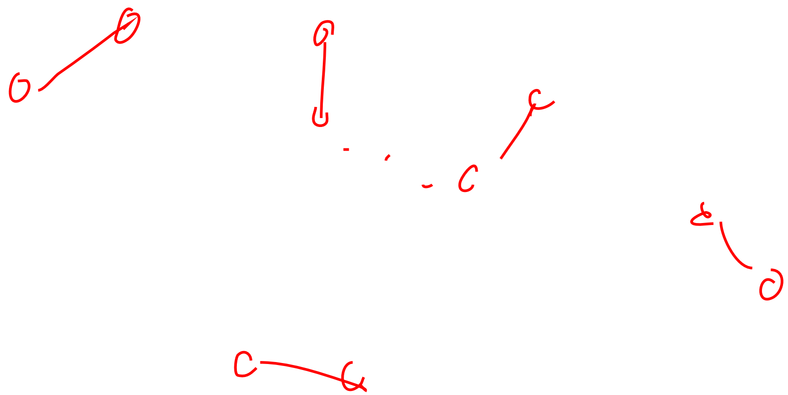
- Prim ✓
- Kruskal ✓

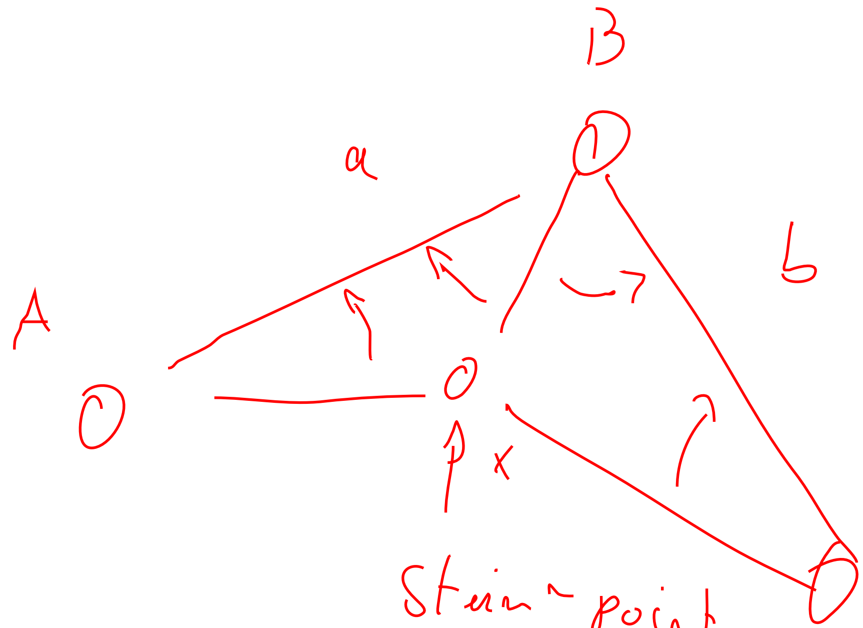
Dijkstra

→ shortest path

Kruskal

- Sort all edges according to weight
- Insert edges into MST if it does not produce a cycle



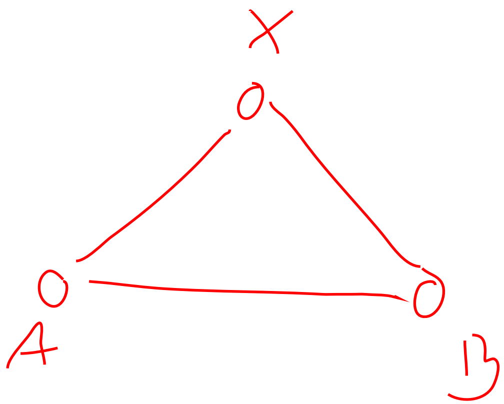


Steiner point

$x \geq a$

$x \geq b$

$2x \geq \boxed{a + b}$



$|A, x| + |x, B| \geq |A, B|$?

Triangle inequality $\Longleftrightarrow$ metric